

Dutch cuts ahead

The new government of the Netherlands should reverse rather than exacerbate its declining support for basic science — a message reinforced by the OECD.

As science policy makers in the Northern Hemisphere head for the summer beaches, few are likely to take with them a weighty report, *Technology, Productivity and Job Creation: Best Policy Practices*, recently published by the Organisation for Economic Co-operation and Development (OECD) in Paris. But the report's message should be remembered on their return. For it is a timely reminder that countries that have developed a healthy knowledge-based economy are those that have combined a vibrant science base with the broader reforms needed to ensure that both research funds and scientific results are used effectively.

The Dutch coalition government of left, centre and right, re-elected to power in May, has secured much of the second, but is in danger of losing sight of the first. In its four previous years in office, the government achieved admirable economic results, reducing the public deficit substantially while holding inflation at 2 per cent and unemployment at 5.5 per cent. Partly as a result, the economy is growing at 3.5 per cent. Its record on research gives less to smile about, however, and the government is now about to pass up an opportunity to use the economic windfalls of the past four years to reverse a sharp fall in research funding over the past decade.

On the contrary, the situation is about to get much worse, with significant cuts planned in science across the research organizations and universities (see page 405). Worryingly, the government seems to have decided that investment should go rather to a handful of applied research and technological infrastructure projects. In

themselves, such projects may be worthwhile. The mistake would be to fund them at the expense of fundamental research. The government has repeatedly declared a commitment to science, and has been hyperactive in its reform of the research system to improve cost-effectiveness, concentrate resources and increase competition among researchers. But with the forthcoming cuts, the plethora of reforms will be seen as a rearrangement of deck-chairs on a sinking ship.

It could be that negotiations between coalition partners have, in effect, elbowed science aside. But Jo Ritzen, the outgoing science minister and a former professor at Erasmus University, must share some of the blame: the record of the past few years alone testifies to a lack of political clout. It can only be hoped that his successor — expected to be the Liberal Party member Louk Hermans — will be more successful in mobilizing support and significantly enhancing the government's perception of science's importance.

The Netherlands is one of only a handful of OECD countries identified in that organization's report as demonstrating 'best policy practice' in its handling of the science base. Some recent changes, such as the regrouping of university research schools, have brought their own discomforts and problems. But the general thrust of the reforms has been positive. It would be a disaster for Dutch science — and, in the long term, for its economy — if they were now to be undermined by ill-considered political penny-pinching. □

A word in your ear

As the conference season approaches, it is time to reflect on the negative effects of ill-prepared presentations.

Conference presentations that leave an audience dumbfounded, confused, exasperated or just plain bored are, unfortunately, a fact of life. A talent for conveying information orally and for well-judged visual presentation is not bestowed on all, while some scientific topics struggle more than others under the necessary weight of details that are as essential as they are overwhelming or intricate. Harder to accept, however, is that some individuals can sustain a substantial career in research without mastering even the basics of public presentation — and continue to ply the conference circuit regardless. Communication skills seem to count for little in comparison with reputation when it comes to invitations to speak.

This is surely disrespectful to audiences that have, after all, often paid considerable registration fees and travelling costs for the privilege of listening. A lack of clarity in the published literature is frequently bemoaned; but at least a reader has the opportunity to pore over a printed text as long as patience allows, slowly dismantling difficult sentences or equations, and digesting complex data. An oral presentation, in contrast, is a one-shot affair. When slides or overheads stuffed with detail are flashed up for barely an instant, when objectives are not clearly stated at the outset, or when highly specialized concepts are assumed to be general knowledge, one might as well have stayed at home and waited

for the paper to appear in print.

Such frustrations are a symptom of the low importance that researchers often attach to the clear transmission of ideas, much as excellence in teaching counts for little when it comes to seeking academic tenure. For oral presentation is, in the end, a form of theatre (if all too often of a breed that Ionesco would have relished). The appropriate skills do not necessarily fall into place as a matter of course; they need to be nurtured (see, for example, the excellent *Public Speaking for Scientists and Engineers* by P. Kenny; IOP Publishing via www.iop.org). This does not imply that excellent presentations need to be flashy or technologically accomplished performances. Nor need it place at a disadvantage those for whom English is not their first language: clear organization and visual presentation do not depend on that.

Neither is it simply a matter of teaching young researchers the right technique; indeed, presentations by newcomers are usually commendably clear precisely because inexperience stimulates more conscientious preparation. Rather, conference organizers should think twice before inflicting on their attendees speakers who are known to show little sympathy for the needs of a broad audience. Better, surely, to interpret their ineptness as a sign of discomfort and to release them from the obligation for their sake as well as for that of everybody else. □



Dutch universities and research laboratories set for major cuts

[PARIS] The Dutch government is about to announce deep cuts across its scientific research organizations and universities. The proposal has already been accepted in principle by the Liberal and Labour parties that make up the new coalition government.

The Liberal/Labour and Democrat alliance that has ruled the country for the past four years re-emerged victorious in the general election on 6 May, with the Labour party led by prime minister Wim Kok polling 29 per cent and the Liberals 25 per cent. After two months of negotiations, the shape of the new ruling coalition and its programme is finally beginning to emerge.

An announcement of the cuts — which are estimated to be around 4 per cent before inflation annually until 2002 — is expected to follow the appointment of a replacement for the current science minister, Jo Ritzen.

The cuts will hit universities and research organizations hard. Their budgets are already stretched because of a recent decline in total research spending, which fell from 2.3 per cent of gross domestic product in 1987 to 1.94 per cent in 1997.

"It is disappointing that, whereas the new government's programmes frequently mention knowledge and technology, additional support has not materialized," complains one ministry official. "Recent cuts have already meant that there is no money to finance some projects needed to bring laboratories up to international levels."

Government spending on research this year totals about Dfl5.5 billion (US\$2.7 billion), with about Dfl3.45 billion of this coming from the Ministry of Education, Culture and Science. The ministry has a budget of Dfl38.9 billion, with Dfl21.7 billion spent on schools and Dfl7.9 billion on higher education. The universities receive Dfl4 billion, and spend Dfl2.4 million on research.

According to the new government's programme, the ministry's overall budget would increase by Dfl0.8 billion a year to 2002. But the extra money will be spent mainly on reducing the size of classes in primary schools. There will also be a substantial increase in funding for information technology use in schools.

Research has fallen victim to the government's determination to curb public expenditure to reduce the budget deficit and to cut taxes. The proposed cuts would have an impact on almost every area of science spending. Universities will suffer cuts of around 3 per cent annually, with more than half the cuts (Dfl75 million) falling on their research funding.

The science ministry's Dfl1.3 billion budget for the country's research organizations, which include bodies such as the national basic research agency, the NWO, will be reduced by around 4 per cent (Dfl45 million) a year.

One result will be increasing pressure on the Netherlands' contributions to international 'big science' projects. These currently account for around 10 per cent of the budget of the research organizations.

The budgetary decisions indicate a significant shift in the government's research efforts away from universities and research agencies towards specific state-led projects geared to socioeconomic goals. The new government programme includes plans to spend on such goals a sum almost equivalent to the proposed cuts in research. These new projects have been drafted by an Interdepartmental Committee for Economic Structure, dominated by the Ministry of Economic Affairs.

Apart from the boost for information technology in schools, projects earmarked for massive investment include the 'Watergraafsmeer' project to create a life and information sciences centre and business park at the University of Amsterdam; the 'Delft Cluster project' to boost research in civil engineering and transport at engineering institutes in Delft, Delft University of Technology and at TNO, the national applied research organization; and the 'Gigaport project' aimed at building a sophisticated information technology infrastructure.

A similar philosophy has guided the creation of a 'superleague' scheme, under which up to ten research schools will be chosen from the 100 or so centres of international excellence. They will be given half the Dfl200

million annual funding earmarked in university budgets for extra support for the research schools.

The research schools were set up five years ago to sharpen the focus of research at the 13 research universities, by regrouping teams in particular disciplines. The reforms were aimed at concentrating resources on the best groups, increasing competition, and boosting the visibility of the country's research abroad. But these goals have been criticized by some researchers who argue that selection is better carried out at the level of individual groups (see *Nature* **385**, 758; 1997).

The government this month selected the first six centres of excellence, each combining departments from several universities. These will receive a total of Dfl245 million over the next five years. The winning schools are in astrophysics, photonics, biomedical genetics, materials, Earth sciences and the chemistry of catalytic converters.

Some are concerned that the cuts could be counterproductive. According to one ministry official, their scale is likely to hamper plans by the government to radically reform the research system.

Under these plans, approved by parliament last April, the NWO would be transformed from a research agency with a dozen of its own institutes into a research council restricted to setting research strategy, and taking over from universities responsibility for distributing up to 25 per cent of grants to university research groups (see *Nature* **390**, 9; 1997). The NWO's institutes, such as the agency of Fundamental Research on Matter, would be merged with institutes of the Royal Netherlands Academy of Science to form a new body under the auspices of the academy.

Declan Butler

Hubble homes in on the birth of bright stars

[LONDON] The Hubble Space Telescope has been used to record a group of 50 young, ultra-bright stars 200,000 light years away in the Small Magellanic Cloud (SMC). They are believed to be the youngest massive stars ever seen in the SMC. Each star is around 300,000 times brighter than the Sun.

The discovery (right) will sharpen the picture astronomers have of how stars formed in distant galaxies after the Big Bang, says Mohammad Heydari-Malayeri of the Paris Observatory, who heads the team of astronomers that made the discovery. The SMC stars, like those before them, are made up mostly of hydrogen and helium. □



M. HEYDARI-MALAYERI & NASA/ESA

China and Taiwan edge a shade closer

[LONDON] A 'private' visit to Taiwan by China's science and technology minister, chemist Zhu Lilan, appears to have helped pave the way for enhanced scientific and technological collaboration between the two long-standing political rivals.

Zhu attended a seminar and spoke with senior academics and industrialists. But at the end of her 10-day visit, she was reluctant to commit herself to a proposal from Hwang Jenn-Tai, chairman of Taiwan's National Science Council — and as such its top science official — that the two sides should sign an agreement setting out guidelines for future substantial cooperation.

Zhu, a prominent polymer chemist who was appointed to her current post earlier this year (see *Nature* 392, 528; 1998), is the first communist Chinese cabinet minister to have visited Taipei since nationalist-ruled Taiwan broke away from the People's Republic of China in 1949. Zhu met leading Taiwanese industrialists, including Morris Chang, one of the founders of Taiwan's semiconductor industry, and Y. C. Wang, prominent in the plastics industry. Other visits included Acer Computer, the National Chiao Tung University and the National Tsinghua University.

According to a report from the Central News Agency in Taipei, Zhu said after a seminar held at Taiwan's Industrial Technology Research Institute (ITRI) that technology on

mainland China appeared to lag behind that in Taiwan by about 10 years. She added that bilateral exchanges, already growing strongly at an unofficial level, should be strengthened.

At a later press conference, she said she could see no major obstacles to increasing scientific and technological cooperation across the Taiwan Strait, scene of major confrontations in the past — most recently in March 1996, when China held naval exercises in the run-up to Taiwan's presidential election.

Hwang said at the end of Zhu's visit that representatives from the two sides who attended the ITRI seminar reached a consensus on the desirability for greater collaboration. He also expressed optimism that both countries will be able to reach an agreement, perhaps a memorandum of understanding, setting out parameters enabling this to happen.

According to the press agency, Hwang said that he was not disappointed that Zhu had not made a formal response to his proposal, as it was the first time that such a proposal had been put forward, and that, formally, Zhu was visiting in a private capacity. Hwang says that important issues needing to be covered include intellectual property rights, administrative support and the respective roles of industries in the two countries in the global market place. "An agreement addressed to these issues, among others, is crucial," he says.

Zhu, who studied organic chemistry in

Russia in the 1950s and Germany in the 1970s, and remains a professor at Peking University, had said on her arrival in Taiwan that she hoped the science and technology communities from both countries would cooperate with "clear paths and targets", rather than loose verbal promises.

The visit was not without its tensions. At one point Liu Zhentao, director of the Cross-Strait Economic and Scientific Exchanges Promotion Centre in Beijing, who headed a 90-member delegation of Chinese industrialists and academics to the ITRI seminar, complained that Taiwan had imposed "too many restrictions" on cross-strait scientific and technological exchanges. In particular, he claimed that China had been forced to cancel plans to display examples of its latest development in biotechnology and in agricultural and defence technologies.

During her visit, Zhu is reported to have said at a dinner with top Taiwanese industrialists that it was impossible to predict whether continuing political differences between the two countries would get in the way of closer collaboration in science and technology. But she expressed confidence that this would not happen. "People who study science are more pragmatic [than politicians]," she said. "This [cooperation] is where our mutual interests lie, and I believe we can definitely push it forward."

David Dickson

New Zealand government backtracks on funding promises

[DUNEDIN] Following a significant cut by the New Zealand government last week in its support for universities, alarm is growing among researchers over the emergence of a radical effort to change the allocation of shrinking funds.

Only two months after a budget that forecast growth, the country is slipping into recession with the coalition government cutting NZ\$316 million (US\$164 million) of public spending to counter the effects of the financial crisis in Asia.

The university sector was the hardest hit, with a cut in the support of student fees of NZ\$50 million. The Treasurer, Winston Peters, claimed that universities could absorb the cut by becoming "more efficient and transparent in administration".

The move breaches a government commitment to 'ring fence' funding for universities by paying a minimum of 75 per cent of the total costs, the remainder to be made up by students' fees. The government share has now been reduced to 72.4 per cent.

Concern over funding is central to a campaign by the Association of University Staff and others to reverse policy decisions



Peters: unrepentant.

believed to have been drafted for a White Paper on tertiary education. These include plans to split teaching and research functions (see *Nature* 392, 320; 1998).

One point at issue is whether the government will direct grants towards short-term applications, and whether the funds will be divided according to peer review or a goal-oriented 'business model'.

"The government simply refuses to listen to and respond to our views," says Kelly Duncan, dean of science at Canterbury University. "Any organizational system should reflect realities, and our present system does not. It is a narrow commercial model, so it constrains, misdirects, inhibits and wastes."

The Ministry of Research, Science and Technology is pushing for research to be reprioritized through a 'Foresight' process

(see *Nature* 391, 426; 1998). This has come under severe attack from academics, who are mystified at a language that includes the terms 'Foresight driven sector clusters', 'proxy purchase' and 'provider capture'. But political pragmatists such as Peter Gluckman, dean of the Auckland Medical Schools, say universities cannot risk losing the little funding they already receive, and have no alternative but to "play the game" in the terms of Foresight and the White Paper.

Meanwhile, the country's nine Crown Research Institutes (CRIs) have had to lay off 72 staff owing to a drop in funding, and are objecting to "unfair" competition from universities. The Association of CRIs says the lay-offs are counterbalanced by increases in staff from other sources of support.

After analysing funding trends, which show government targets and real funding moving further apart each year, the 18 council members of the New Zealand Association of Scientists jointly conclude: "Reprioritizing science on a background of a static or falling science budget will only increase the present fears of the science community."

Peter Pockley

Ministers block disposal of oil rigs at sea

[LONDON] Europe's environment ministers last week ended three years of public controversy about the fate of disused oil rigs in the northeast Atlantic ocean. They decided that most will have to be dismantled onshore, rather than disposed of in the sea.

The decision was made at a meeting of countries belonging to the Oslo-Paris (Ospar) convention on protecting the marine environment. Ministers attending the meeting in Sintra, Portugal, also agreed to reduce radioactivity in emissions from nuclear plants to "close to zero" by 2020 (see below).

Greenpeace welcomed the decisions, as did virtually all environmentalist groups. Oil companies, on the other hand, were disappointed. The UK Offshore Operators Association said the decision to outlaw deep-sea disposal of oil and gas rigs appeared to have been based on political expediency, not on science.

But the industry took some consolation in a caveat to the decision: that oil installations heavier than 10,000 tonnes did not have to be totally taken apart. Instead, an installation's heavy concrete base — the most expensive part to dismantle — would be left on the sea bed for the time being.

The eventual fate of 41 such structures in the northeast Atlantic would then be decided on a 'case by case' basis. This is an important caveat, as it allows the oil industry to divert the estimated £1 billion cost of removing 41 stumps into a fund to address what they consider to be more pressing environmental concerns.

If the stumps are allowed to remain where

they are, some oil executives are privately prepared to consider putting the money saved into a trust, known informally as the 'green super-fund'.

"This is the opportunity of a lifetime," says the head of decommissioning of a major oil company. "We're going to spend the money anyway, and we might as well spend it on something more environmentally worthwhile than removing some concrete from the bottom of the sea."

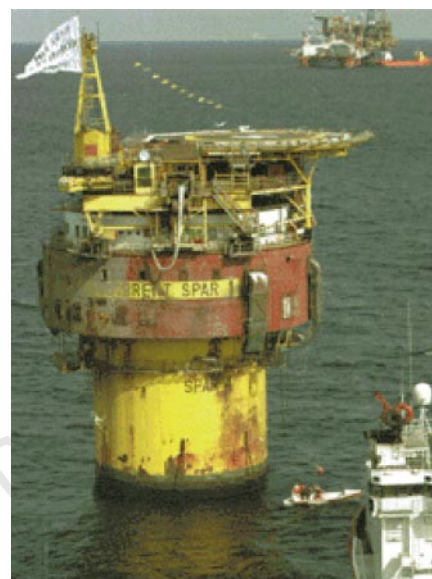
Environmentalists are divided on whether to support the idea of a super-fund. Friends of the Earth is the only one of the big three groups to come out in favour of it. Greenpeace and the World Wide Fund for Nature (WWF) are opposed to being so closely associated with oil industry funds.

Sian Pullen, head of WWF's marine unit, says that exposed stumps could endanger the safety of fishermen working nearby. She also says that the oil industry has a responsibility to clear its sites from the sea.

But David Bellamy, president of the Conservation Foundation in London, is enthusiastic about the super-fund. He describes the idea as "an historic opportunity" to invest in environmental protection, which should not be lost through adherence to dogma.

Pete Wilkinson, a former director of Greenpeace, agrees. "We can do some real good here. Let's not get hung up about peripheral issues. It's nonsense to say we can't take oil industry money, when Greenpeace puts money in an investment fund that includes oil companies."

The origins of the ban on deep-sea disposal lie partly in a 1995 dispute between



Brent Spar: highlighted the disposal debate.

Greenpeace and the Shell oil company over Shell's plans to sink Brent Spar, a disused oil storage platform in the north Atlantic (see *Nature* 375, 708 & 376, 378; 1995).

Greenpeace charged that deep-sea disposal was harmful to the marine environment and tried physically to stop the disposal. A vigorous public controversy ensued, which forced Shell to change its mind. The company later embarked on a lengthy — though more transparent — process to settle on its alternative plan of re-using parts of the Brent Spar as a quay extension in Norway.

Scientists remain bitter about the affair, which to many was a triumph of emotion over reason, and a clear example, they argued, of the need for more public "understanding" of science.

In its first report, a year after the controversy, an expert group set up by the British government said that the global environmental impact of deep-sea disposal for Brent Spar would be "very small, roughly equivalent to the impacts associated with the wreckage of a fairly large ship".

And in its latest report, published last month, the group concludes that, from an environmental perspective, there is little to choose between deep-sea disposal and the new alternative.

Mike Cowling, professor of Marine Technology at the University of Glasgow and a member of the group, says that Brent Spar should not be used as a precedent for decisions on the future disposal of the majority of offshore oil and gas rigs.

Brent Spar was about the environmental impact of spilling a small amount of oil into the sea, he says, whereas most conventional oil rigs do not store oil.

Meeting agrees cuts to radioactive emissions

[LONDON] Europe's nuclear industry will have to make "substantial reductions" to emissions of radioactive substances by the year 2000 following a decision reached by the continent's environment ministers at last week's meeting of the so-called Ospar convention (see above) in Sintra, Portugal.

The legally binding agreement requires nuclear operators to reduce radioactive discharges to "close to zero" before 2020. In addition, the chemical industry will not be allowed to emit "hazardous substances" into the environment unless they have been demonstrated and certified to be safe.

The decisions have been hailed as historic by environmentalists. Helen Wallace, senior scientist at Greenpeace, says the agreement is a rare example of the precautionary principle at work. "The burden of proof [that discharges are safe] is now on industry, and not on the consumer."

Some environmentalists predicted that the agreement could force the closure of facilities such as Britain's nuclear waste reprocessing plant in Sellafield. But the ambiguous phrase "close to zero", and the proviso that reductions will not be mandatory until they are technically feasible, are likely

to ensure their survival.

Much of the agreement was influenced by John Prescott, Britain's deputy prime minister and secretary of state for the environment, who again demonstrated his effectiveness as an international political broker. Prescott's deal-making skills helped to forge agreement between Europe and the United States in December's climate talks in Kyoto, Japan.

In the short term, facilities such as Sellafield and the French reprocessing facilities at Cap de la Hague on the English Channel will have to divert radioactive discharges into dry storage instead of into the sea.

E.M.

Cloned mice fail to rekindle ethics debate

[WASHINGTON] Just over a year ago, the US Congress was in ferment over proposals to enact legislation outlawing human cloning, following the cloning of Dolly the sheep. But little has happened since, and legislators seem unlikely to take immediate action in the wake of the announcement last week that Hawaiian scientists have made dozens of mouse clones, a species previously thought to pose prohibitive biological obstacles (see *Nature* 394, 369; 1998).

The report, which drew predictions that human cloning may be much nearer than expected, has provoked muted reaction from Congress, the White House and the National Bioethics Advisory Commission (NBAC), which had all previously called for human cloning to be either outlawed or subject to a legislated moratorium.

"The president's position has not changed," says Estela Mendoza, a White House spokeswoman. "He still believes it's morally unacceptable for anyone in either the public or private sector to attempt to produce a human being using somatic cell nuclear transfer technology."

Harold Shapiro, president of Princeton University and NBAC chairman, says the work presents "no reason to be alarmed", but should spur intense ethical discussions.

In Congress, a tight legislative schedule before October's recess, combined with a split between libertarian and socially conservative Republicans on the advisability of outlawing cloning, is likely to forestall any imminent action.

This is good news for many biomedical research advocates, who have backed volun-



Double or quit? Not everyone shares Dr Ryuzo Yanagimachi's delight in his cloning success.

tary moratoria on cloning, fearful that a law could unduly restrict research. Shapiro agrees: "We're much better off without legislation than with bad legislation."

No law currently prohibits privately financed human cloning in the United States, although federal funding is prohibited under a broader ban on government funding for human embryo research.

There is a deep division between moderate and conservative Republicans in the Senate, the former not wanting to constrain research, the latter demanding a criminal ban on the production of cloned embryos, whether for research or implantation.

Bill Frist (Republican, Tennessee), a physician who earlier this year co-authored a

bill that would have criminalized human cloning, whether or not cloned embryos are implanted, describes the news of the cloned mice as "predictable" but says "this just probably cut by 50 per cent the length of time to clone a human being".

Asked at a Senate news conference whether Congress should act on his bill — which was shelved in February after lobbying by biomedical research advocates and fertility physicians (see *Nature* 391, 730, 1998) — he called instead for Congress to pass another bill he has authored.

This calls only for the establishment of a national bioethics commission to discuss cloning among other issues. It would allow Congress to act intelligently and "not be forced to react and respond to the latest news event".

NBAC last year recommended a three-to-five-year moratorium on human cloning, basing its arguments on safety considerations (see *Nature* 387, 644; 1997). This drew complaints from conservative Republicans, who said a permanent ban was needed because moral considerations make cloning unacceptable.

Edward Kennedy (Democrat, Massachusetts), co-author with Dianne Feinstein (Democrat, California) of another Senate bill, says the news changes nothing in terms of the approach to policy. Kennedy's bill would impose a ten-year moratorium on human cloning, but not on the production of cloned embryos that are not implanted. As such, it "represents the best public policy", Kennedy said last week.

The leading author of anti-cloning legis-

Japanese fear that new publicity rules could hinder their research

[TOKYO] Japanese scientists are concerned that proposed regulations on cloning could jeopardize the international impact of their work.

An interim report released in June by the Science and Technology Council expresses support for the advancement of cloning technology (see *Nature* 393, 726; 1998). But as well as calling for the strict regulation of human cloning, it requires information about all mammalian cloning research to be made public before the animal is born.

The report, to be finalized by autumn, says researchers are not expected to release information that could undermine their intellectual property rights. But it does not specify precisely when

they are to release their information. Scientists will be expected to report their plans in advance to the government, which will make a public announcement to stimulate awareness and understanding of cloning technologies.

Researchers carrying out projects funded by the Ministry of Agriculture, Fisheries and Forestry (MAFF) will be required to announce the contents of their research activities a hundred days after implantation.

But several researchers are concerned that these regulations would make it impossible for them to conceal the birth of cloned animals until publication in scientific journals. They see this as a significant disadvantage,

as researchers outside Japan are not subject to such restrictions.

Yukio Tsunoda, a professor in mammalian embryology at Kinki University who was responsible for the cloning of twin calves earlier this month (see *Nature* 394, 114; 1998), says he feels under pressure because the paper he is writing about it will be published months after the animals were born.

The calves, produced by scientists at Kinki University and Ishikawa Prefecture Livestock Research Centre, are claimed to be the second large mammals cloned from adult somatic cells. The team used cells from the oviducts of an adult cow.

Apart from the fact that the project was funded by MAFF,

Tsunoda says his decision to announce the births of the calves reflects the government's call for the disclosure of scientific information on cloning.

Although news of the cloned calves was welcomed by the scientific community in Japan, the lack of detailed information gave rise to some scepticism among overseas researchers. One cloning expert says he still has no idea of the techniques or type of cells used in the work.

The twin calves, named Noto and Kaga, will soon be joined by many others. MAFF has recently revealed that a total of 29 cloned calves — including eight from Tsunoda's research groups — are expected to be born by the end of the year.

Asako Saegusa

lation in the House, Vern Ehlers (Republican, Michigan), says temporal and political constraints will almost certainly block action this year. He says "there are not enough votes" to move his bill out of the health and environment subcommittee of the House Commerce Committee where it is stalled. His bill would ban federal funding for producing cloned human embryos, whether or not they are implanted.

The Biotechnology Industry Organization (BIO), which backs a voluntary moratorium, says it has not changed its position. "BIO stands against the use of this technology to clone human beings," said Carl Feldbaum, BIO's president, in a statement. It says the Food and Drug Administration (FDA) can block cloning under existing federal regulations, a position the FDA itself asserted in January after Richard Seed, a Chicago physician, vowed to clone a human.

"If you can clone from mice, humans should be easier," says Lee Silver, professor of genetics at Princeton University and author of *Remaking Eden: Cloning and Beyond in a Brave New World*. "It's going to be used by [a small percentage of infertile] people to have biological children, period," Silver says. "That's the only way it's ever going to be used. And I have no problem with that."

But Robert Edwards, the British pioneer with the surgeon Patrick Steptoe of *in vitro* fertilization, says infertility does not justify cloning. Edwards, who last week endorsed the idea that the test-tube embryos of infertile couples be cloned in difficult cases to improve the chances of implantation, says that the use of cloning in infertility treatment should stop there.

"Many people would say it's the thin end of the wedge," Edwards concedes. But, he adds, "nobody's worth cloning... I can't think of anybody even to mention apart from the Lord Himself."

Meredith Wadman

Patent clash looming over cloning techniques?

[PARIS] A clash over patent rights to the technology of cloning by nuclear transfer seems on the cards, following an application for patents by the University of Hawaii and ProBio America Inc, a Honolulu-based biotechnology company, on the technology behind last week's report on the reproducible cloning of mice.

The Roslin Institute in Scotland, which cloned the sheep Dolly two years ago and has two broad patent applications pending covering cloning techniques, says that if it feels that the new claims infringe these it will challenge them.

"We will defend our intellectual property position vigorously," says Harry Griffin, assistant director of the Roslin Institute. He emphasized that the precise claims in the new patent application are not yet clear.

To produce Dolly, an adult cell was fused with an enucleated oocyte by electrofusion and the reconstructed embryo was simultaneously activated. The Honolulu group says the novelty of its method is in directly microinjecting the nucleus into an enucleated egg and activating the egg later – a step claimed to partly explain the improved success rate (see

Nature 394, 369; 1998).

But Griffin says: "I don't think that injection of a nucleus rather than cell fusion, or the delay in activation, is novel." He adds that the Roslin patent applications cover "all quiescent cells", so include the cumulus cells used by the Hawaiian researchers. "We believe we have a very strong patent position," says Griffin.

No details of the Hawaiian application have been made public. The University of Hawaii declines to comment in detail on its application, saying that this is "proprietary information".

Andrew Sheard, a patent attorney at London-based Kilburn & Strode, which is handling the Dolly patents, says: "It seems to us that the Hawaiian group have made a refinement of the basic technology." He says the Roslin patents cover cloning broadly, in that they claim "a method of reconstituting the embryo comprising transferring the nucleus of a quiescent donor cell into a suitable recipient cell".

Meanwhile, UK company Protein Pharmaceuticals Limited (PPL) has moved swiftly to negotiate a licence on the ProBio America technology for research into the cloning of pigs and other farm

animals. The company has already agreed to take a licence on cloning technology from the Roslin Institute. But this only covers "production of proteins for pharmaceutical and nutraceutical use in the milk of ruminant livestock or rabbits".

Roslin has exclusively licensed rights for all other biomedical applications of genetically modified livestock, including use of pig organs as human transplants, to Roslin Biomed, a company set up with £6 million (\$9.9 million) from the venture capital company 3i.

PPL, which has a substantial research effort in xenotransplants – a potential market of \$6 billion (see *Nature* 391, 315; 1998) – sees the Honolulu patent as a route into pig cloning. "The potential of cloning for breeding animals, and in particular pigs, for xenotransplantation, is certainly the major reason for interest," says Ron James, chief executive officer of PPL.

PPL has no access to the Roslin technique for xenotransplantation research, and James argues that the Hawaiian technique seems more promising for pigs.

"Ultimately, he says, one may need both the Roslin techniques and those of the Hawaiian researchers."

Declan Butler

Mother bears could help save giant panda

[TOKYO] Global efforts at cloning do not stop at familiar animals such as sheep, cows and mice. Scientists in China are planning to attempt to clone the country's national symbol, the giant panda, in a desperate bid to save the animal from extinction. But other researchers are sceptical about their chances of success.

Last week, scientists from China's National Academy of Sciences announced a project to clone the giant panda by 2003. They are hoping to transfer the nucleus from an adult giant panda cell into the egg of another species, perhaps a black bear. The hybrid egg would be induced to differentiate before being implanted into the uterus of a foster mother of the second species.

The researchers say the advantage of such a technique is that only a few panda

cells would be needed to produce an embryo. Such 'trans-species' cloning would also avoid using panda eggs, which — with only 1,000 pandas left in the world — are rare and difficult to harvest.

Chen Dayuan, the leader of the project and a professor of zoology at the academy, expressed his confidence, during an interview with China Central Television, that cloning could become an effective method to breed pandas, and that Chinese scientists would be the first to succeed with such an experiment.

But international experts in cloning have doubts about the chances of success, pointing out that cross-species cloning is unlikely to work unless the eggs and the genetic material are taken from closely related species.

A.S.



Making babies: artificial insemination attempt in 1996. Could cloning save Jia Jia's species?

AP/MAX VISION

Alarm in US over database antipiracy bill

[WASHINGTON] Legislation aimed at strengthening the intellectual property protection of commercial databases is moving quickly through the US Congress, alarming many science advocates who say it will have a chilling effect on the free flow of information among academic researchers.

The Collections of Information Antipiracy Act (HR 2652), sponsored by Howard Coble (Republican, North Carolina), passed the House of Representatives with little debate in May. A Senate counterpart, S 2291, was introduced by Rod Grams (Republican, Minnesota) this month, and is scheduled for hearings in the autumn.

The proposed legislation would prohibit the extraction or commercial use of "all or a substantial part... of a collection of information gathered, organized, or maintained by another person through the investment of substantial monetary or other resources".

The bill is designed to protect large commercial databases from electronic piracy. It is backed by a powerful group of publishers, including Reed Elsevier (publisher of the Lexis Nexis database and Elsevier science journals) and Thomson Corporation, parent company of the West Group, which owns the Westlaw online legal research service and is based in Grams' home state of Minnesota.

But the bill is opposed by science and library associations, including the American Association for the Advancement of Science (AAAS) and the Association of American Universities (AAU).

They warn that the protection offered to commercial database owners is too sweeping, and could have serious unintended consequences for the research community. Scientists could end up paying for access to their own data after they have been repackaged



by a commercial database owner.

Researchers who adapt, modify and share data from a wide variety of sources could become paralysed by the fear of lawsuits. "The potential for litigation is phenomenal," says Debra Stewart, vice-provost at the North Carolina State University, who testified against the bill on behalf of the AAU earlier this year.

The legislation would be "a mess, a disaster", warns Mark Frankel of the AAAS, adding that it could lead to a fundamental shift in the culture of science. Entrepreneurial universities would come to view research data as assets to be hoarded and marketed, rather than as knowledge to be freely shared, he says.

The US science community fought this battle as recently as 1996, when it successfully lobbied to block a proposed treaty on database protection being considered by the World Intellectual Property Organization (see *Nature* 384, 299; 1996).

But the ease with which the Coble bill moved through the House — despite letters of opposition and testimony from science and library associations — suggests that the

community's concerns are not taken seriously. "We were very ineffective at making our case, even with people [like Coble] who are our friends," says Stewart.

She and others argue that, although the proposed legislation appears to exempt researchers, news organizations and other non-profit users from restrictions on data use, the bill's language is so vague that it offers little protection for scientists. "A real exemption for universities could go a long way" towards easing the research community's fears, says Stewart.

Coble accuses scientists of worrying too much. But George Brown of California, the senior Democrat on the House Science Committee, answered that "[scientists] may be worried because they do not understand the bill," and called for Congress to go slower in passing the antipiracy legislation. "I would like to alert a broader audience to the fact that there could be some flaws."

The White House has been silent on the issue, but is expected to weigh in with its opinion this week, in the form of letters to key senators. The Office of Science and Technology Policy has been formulating the administration's position, and President Bill Clinton's National Economic Council has discussed it.

The Coble bill's quick passage through the House was aided by the work of high-powered lobbyists, including Laura Tyson, once Clinton's chief economic adviser and now a consultant to Reed Elsevier and Thomson. The Grams bill is expected to move more slowly through the Senate, giving opponents hope that it can be modified or thrown out.

Political divisions over the issue are not clear cut. The Information Technology Association of America (ITAA), which opposes the legislation, counts among its members large database owners such as Dun & Bradstreet, as well as companies such as Microsoft, Netscape, MCI and America Online.

The ITAA argues that concerns about piracy should be addressed with a more circumscribed amendment to existing copyright law. One critic of the Coble-Grams bill says it tries to "kill a small problem with an assault rifle".

Several groups, including the ITAA and the Institute of Electrical and Electronics Engineers (IEEE), have suggested amendments to the bill to make it less onerous to researchers. Instead of prohibiting the "use" and "extraction" of data, the IEEE suggests a ban on "unauthorized selling and distribution of data", for example.

The ITAA and its allies hope this kind of modification will turn the antipiracy law into something both sides can live with. "We obviously want greater database protection," says the association's general counsel, Marc Pearl. "We just don't think this [bill] is the best approach."

Tony Reichhardt

Telescopes track down lost spacecraft

[MUNICH] Ground-based radiotelescopes have located the \$1 billion US-European solar space mission SOHO, which slipped out of radio contact at the end of June.

Until now, all attempts to find the spacecraft have failed. But signals transmitted last week by the US National Astronomy and Ionosphere Center in Arecibo, Puerto Rico — the location for the film *Contact* — were reflected back to Earth and picked up by the 70-metre dish of NASA's Deep Space Network in Goldstone, USA.

The batteries of SOHO (the Solar and Heliospheric Observatory) appear to have been drained during the craft's unsuccessful attempt to maintain its correct orientation towards the Sun, after this had been lost during maintenance (see *Nature* 394, 5;

1998). In its current orientation, edge-on to the Sun, the spacecraft cannot use solar energy to recharge the batteries.

But engineers at NASA and the European Space Agency have confirmed that SOHO is spinning slowly, at exactly the rate they had predicted, suggesting that little structural damage has occurred. They have also confirmed that SOHO has not yet shifted from its correct orbit.

Scientists are still hoping that SOHO will, in the course of its current orbit, move into a position where it leans sufficiently towards the Sun to allow recharging of its batteries. This might happen within three months. Once recharged, SOHO could once again respond to radiocommands and transmit data — providing no serious damage has been done.

Alison Abbott

A physicist joins the race for Congress

Colin Macilwain

There are only three scientists in Congress, but Rush Holt, former assistant director of the Princeton Plasma Physics Laboratory, believes he can become the fourth. He comfortably won the primary to become the Democrat nominee in New Jersey's twelfth district.

[FREEHOLD, NEW JERSEY] "People are surprised to find a physicist who is adept at this," says Rush Holt, the former assistant director of the Princeton Plasma Physics Laboratory, as he pauses between a quick canvassing tour of a retirement community and a sparsely attended, open-air press conference on a wet, muggy morning in New Jersey.

Holt left the laboratory, the largest fusion research facility in the United States, last October to run for Congress in New Jersey's twelfth district.

The evidence of Holt's adeptness is his unexpectedly comfortable victory in a primary election in June, when he easily defeated a wealthy rival, Carl Meyer, to become the Democratic nominee for this well-heeled strip of suburban New Jersey.

Now he stands, as his literature puts it, as "teacher, scientist, Democrat" for one of the handful of the 435 congressional districts that political analysts believe may actually change hands in November.

If Holt wins, he would be the fourth PhD scientist in Congress, joining physicist Vernon Ehlers (Republican, Michigan), chemist John Olver (Democrat, Massachusetts) and physiologist Roscoe Bartlett (Republican, Maryland). But, as a one-time arms-control official at the State Department who spent much of the past ten years trying to secure support for fusion research in Washington, he would be the first to arrive at Capitol Hill with extensive science policy experience.

Holt is standing against Republican Michael Pappas, who won the seat in 1996 with only 49 per cent of the vote. This narrow margin of victory, and the fact that Pappas is a 'freshman' member who has only had two years to accumulate goodwill in the district, puts it near the top of the target list of Republican seats that Democrats hope to win in November. They need 11 gains to regain control of the House of Representatives.

Holt's campaign seeks to paint Pappas as an "extremist" who supports policies on issues such as abortion and the environment that are out of line with the district's moderate electorate. Pappas has been walking a fine line to protect himself from this charge. The League of Conservation Voters, for example, gives his environmental voting record the lowest ranking of any New Jersey congress-

man, but the highest ranking of any Republican freshman in the entire House.

Holt concedes that Pappas is not the easiest of targets. "He keeps his head down, but his voting record shouts 'extremist.'" Pappas failed to respond to calls asking him to defend his record.

If elected, Holt does not plan to seek any radical changes in the way Congress allocates money for research. He also says that any attempt to centralize responsibility for science policy, for example in a Department of Science, would be a "disaster". But he adds that he will work to get Congress to respond to global warming and biodiversity conservation, science-driven issues that he claims Pappas has ignored.

At this early stage, the top priority of both candidates is raising money. Holt says he will need \$1 million, but expects Pappas to raise at least \$1.25 million. It was reported in the district that the Democratic Party in Washington wanted Meyer to win the primary, as he had \$1 million of his own to spend, whereas they doubted Holt's ability to raise the cash needed to make the race competitive.

Holt confounded them, raising \$311,000 by the end of June. Much of it has come from scientists, including former colleagues, science policy professionals in Washington and scientists from across the country, including half-a-dozen Nobel laureates.

"These are not traditional donors, but [they] feel deeply about the issues," he says. The traditional contributors — including political action committees run by environmental groups and labour unions — can be expected to pour money into the later stages of the campaign if the seat looks winnable.

The district meanders from the rural county of Hunterdon down through Princeton almost to the Jersey shore, and has no real centre or dominant newspaper.

Television advertising would be inefficient and prohibitively expensive in a district sandwiched between Philadelphia and New York, so the best way to reach voters will be through grass-roots campaigning. Holt was enjoying this last week, greeting voters — mostly Jewish retirees from New York City, who will draw their last breath before voting Republican — at the Covered Bridge retirement community.



Holt: if elected, he would push Congress to act on global warming and biodiversity conservation.

"He has a good chance of winning," says Harry Sayen, a columnist for the *Trenton Times*, who says that moderate Republicans in the Princeton area will desert Pappas. "Republicans voted for him in 1996 without realizing what kind of Republican he was."

Although Holt promotes himself as a scientist, Sayen suggests he keeps quiet about it, even in a district where as many as two thirds of voters had a college education. "You shouldn't push academia too hard, or people think of you as an egghead," he says.

That advice is echoed by Vernon Ehlers, who says he decided not to emphasize his scientific credentials when he first ran for Congress in 1993. But being a scientist brought him a decisive break in that campaign.

In a televised debate, two of his rivals for the Republican nomination were boasting that their backgrounds, as a lawyer and a businessman respectively, would make them effective on Capitol Hill. "I just said that I had checked and there were 181 lawyers and 135 businesspeople in the Congress," Ehlers recalls. "But if I were elected, it would double the number of scientists." He won the primary and the election.

George Brown (Democrat, California), the senior Democrat on the House Science Committee, feels that more scientists and engineers are needed in the legislature. "There's no reward structure for the scientist who goes into politics," says the veteran Congressman. "The pay's not great and they lose their status in their own community."

That doesn't bother Holt, who believes he can win in November even without a national swing to the Democrats. A poll for his campaign says only 30 per cent of district voters want to re-elect Pappas. "The polling data is very good, and I'm finding good vibes in the district. Those who know me best are working hard for me, and that's not always true for politicians." □

Brussels may open Framework projects to eastern countries

[BRUSSELS] The European Commission has proposed that 11 countries that are candidates for European Union (EU) membership should be able to participate fully in the fifth Framework programme of research (FP5), provided they can make a significant financial contribution.

The 11 countries are Bulgaria, Czech Republic, Cyprus, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia. Many will be able to use money from the EU's PHARE programme, set up to assist central eastern European countries, to pay for their contribution to FP5.

The commission's proposal must be approved by the council of ministers before negotiations begin between the commission and individual countries over their financial contributions.

Climate change disasters looming, warns insurer

[MUNICH] Munich Re, the world's largest reinsurance company, has called on governments to sustain their actions against manmade climate change, which it believes is responsible for a global trend towards more frequent and severe storm and flood disasters.

The company said last week that the costs of natural disasters in the first half of 1998 (US\$24 billion) have already exceeded the total 1997 costs. More than 20,000 people have already died in such disasters, compared to an annual average of around 10,000.

Gerhard Berz, head of the company's Earth sciences research group, warned that global warming could lead to a "permanent El Niño".

Berz said: "We cannot afford to wait for definite scientific proof of global warming, and would be better advised to be prepared for the worst."

Russia cuts red tape from research grants

[MOSCOW] Russia's Ministry of Finance has agreed to rescind an instruction under which the Russian Foundation for Fundamental Research had been obliged to plan all spending on grants a year ahead. This has been highly inconvenient for the foundation and for grant recipients, who were only able to use their money if they had been able to foresee the need for it many months before.

The foundation was set up by presidential decree in April 1992. It receives 6 per cent of all federal money allotted to science, with about 16,500 Russian scientists receiving its grants. But its present financial situation is

bad. In the first six months of this year, it only received 161.7 million roubles (US\$27 million) out of the 854 million roubles it had been promised by the government.

NIH defends use of the diet drug fenfluramine

[WASHINGTON] The US National Institutes of Health last week submitted to Congress documents defending the use of the diet drug fenfluramine, which was recalled after it was found to cause heart valve injury in psychiatric experiments on children in New York City. The drug's use is being investigated by Congressman Dan Burton, the chairman of the House Committee on Government Reform and Oversight (see *Nature* 393, 406; 1998.)

"There still are no known serious side-effects associated with the single low-dose, oral administration of the drug used in these studies," wrote Steven Hyman, the director of the National Institute of Mental Health. "At no time has single-dose ingestion of fenfluramine been linked to cardiac valve injury," he added.

Farmer fails in bid to halt trial of modified maize

[LONDON] Britain's Court of Appeal last week rejected a plea from an organic farmer to halt a trial crop of genetically modified maize that is planted close to his own crop of organic sweetcorn in Devon (see *Nature* 394, 212; 1998).

The farmer, Guy Watson, now risks losing his organic certification if cross-pollination occurs between the two crops. However, the judges ruled that this was unlikely because the two crops are two kilometres apart.

But the judges also ruled that the government had acted illegally by not carrying out two pilot trials of the modified maize before the current 'national seed list trial', from which a decision will be taken on whether to allow the maize to be sold commercially.

Schools get poor marks for maths teaching

[LONDON] British schools should devote more time to teaching pure mathematics, even if this means cutting back on statistics and mechanics, according to a report published by the Royal Society last week. It says the level of difficulty, compared with other subjects, of 'advanced' level mathematics and science, taken between the ages of 16 and 19, is "a major concern" and is deterring students from studying these subjects.

The report recommends cutting down on the breadth of topics covered, and

concentrating on developing fluency in a smaller number of key concepts. It also raises concerns about the tendency of some schools to enter pupils in less challenging examinations to raise the number of top grades awarded.

Novartis digs deep for plant genomics institute

[WASHINGTON] The pharmaceutical company Novartis signalled its ambitions in the field of agricultural biotechnology last week by announcing that the Novartis Research Foundation is to invest US\$600 million in plant genomics over the next decade.

Around \$350 million has been reserved for industrial-academic collaborations. The rest will be spent on setting up and running the Novartis Agricultural Discovery Institute in La Jolla, California. Novartis has begun advertising for the institute's expected 200-strong staff of specialists in genomics, plant sciences and related fields. A new building is expected to be completed by next year.

Brazilian to head WHO's tropical disease initiative

[LONDON] Carlos M. Morel, former head of biochemistry and molecular biology at the Oswaldo Cruz Foundation in Rio de Janeiro, Brazil, has been appointed director of the World Health Organization's (WHO) Special Programme for Research and Training in Tropical Diseases (TDR). Morel replaces Tore Godal, who retired recently after 12 years and is currently helping to manage WHO's 'roll back malaria' campaign.

TDR has been strengthened under restructuring plans being implemented by Gro Harlem Brundtland, the new director general of WHO. Its co-sponsoring agencies include the World Bank and the United Nations Development Program.

US 'can afford large cuts to greenhouse gases'

[WASHINGTON] The United States could reduce greenhouse gas emissions by nearly twice the 7 per cent called for in the Kyoto protocol, with net savings to the economy, according to a report released last week by the Union of Concerned Scientists and the Boston-based Tellus Institute.

The report — *A Small Price to Pay: US Action to Curb Global Warming is Feasible and Affordable* — claims that the United States could reduce emissions by 2010 by as much as 13 per cent below 1990 levels, using "domestic actions that will have zero or negative net costs". It analyses recent studies by groups including the Department of Energy, the Office of Technology Assessment and the National Academy of Sciences to arrive at its conclusions.

Japan to the fore in biodiversity initiative

Sir—I read with interest your Briefing “Museum research comes off list of endangered species”, which described the establishment by the Organization for Economic Cooperation and Development of the Global Biodiversity Information Facility (GBIF)¹.

Although efforts by museums to collect data on species are laudable, I believe the emphasis should be put on microorganisms because, to date, fewer than 1% of the world's microorganisms have been identified and described². To ensure a comprehensive collection of data on biological diversity, emphasis should be placed on the taxonomy of microorganisms.

Reference was also made to ‘Species 2000’, which held its first international

meeting in the United Kingdom in 1994. The article did not mention that from the beginning this project was led by a management team of international collaborators³ and the software used was developed and refined in Japan especially for the study of microorganisms.

In the same manner as European and North American museums and botanical gardens have collected data on biodiversity for many years, Asian scientists have been working together to develop a stable collection and data network under the title Asian Network on Microbial Researches, funded by the Japanese Science and Technology Agency⁴. It would seem logical for the GBIF to amalgamate collections in Europe and North America with those in Japan and Asia.

I realize my opinion is slightly biased, but I believe that the data collection system in Japan is cutting-edge technology, and the most appropriate to be used for the GBIF.

By working together, it can be ensured that funding will be efficiently distributed among the biologists most in need and most capable of doing this research on biological diversity.

Junko Shimura

Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa Wako, Saitama 351-0198, Japan

1. *Nature* **394**, 115–119 (1998).

2. Hawksworth, D. L. & Kalin-Arroyo, M. T. in *Global Biodiversity Assessment* (Cambridge, 1995).

3. www.sp2000.org/sp2000bg

4. Shimura, J. et al. *WFCC Newsletter* (in the press). www.jcm.riken.go.jp/WFCCfina

Extrasensory statistics

Sir—In his review¹ of my book *The Conscious Universe*, I. J. Good suggests that something “must be” wrong with my statistical arguments in favour of the reality of psychic phenomena. In fact, his inability to reproduce my estimate that 3,300 unpublished, unsuccessful experiments would be required for each published extrasensory perception (ESP) card experiment, in order to nullify the outcome, follows from a misunderstanding.

Good assumed incorrectly that my words “more than” (in the phrase “more than a billion trillion to one”, referring to the cumulative odds against chance for the 186 experiments listed in ref. 2) could safely be replaced by “approximately equal to”, implying a *P*-value of about 10^{-21} . Instead, the real *P*-value is about $10^{-2,000}$, which, following the application of standard methods³, gives my reported figure.

I am encouraged that a statistician of Good's repute did not discover any genuine flaws in my comprehensive analysis of the empirical evidence for parapsychological phenomena.

Dean Radin

Interval Research Corporation, 1801 Page Mill Road, Building C, Palo Alto, California 94304, USA

I. J. Good replies—Radin's claim of my “misunderstanding” relates to four million guesses of cards made up to the year 1939. In his book (p. 97) he says that the combined result amounts to “odds against chance” of more than a billion trillion (10^{21}). Most readers would take this as implying that the real odds were not more than a trillion trillion (10^{24}), and would be surprised to hear that, by “more than a

billion trillion”, Radin meant more than 10^{100} —increased to $10^{2,000}$ after the book was published (information from Radin; Brian Josephson, personal communication).

Readers should bear in mind that Robert Rosenthal's analysis of the ‘file-drawer effect’⁴ makes no allowance for intellectual, observational or ethical lapses; nor does Radin make any such allowances in his calculation. S. G. Soal, regarded by many as the leader of ESP research for many years, was shown by Betty Markwick⁵ to have almost certainly been a cheat, thereby confirming previous suspicions. Radin ‘abolished’ Soal, in the sense of George Orwell's 1984. The proportion of lapses in parapsychological studies is difficult to estimate, and is made more difficult by ‘abolishments’. An abolishment of an important lapse is another lapse.

I. J. Good

Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0439, USA

1. Good, I. J. *Nature* **389**, 806–807 (1997).

2. Pratt, J. G., Rhine, J. B., Smith, B. M., Stuart, C. E. & Greenwood, J. A. *Extrasensory Perception After 60 Years* (Bruce Humphries, Boston, 1966).

3. Rosenthal, R. *Meta-analytic Procedures for Social Research* 2nd edn (Sage, Newbury Park, 1991).

4. Rosenthal, R. *Psychol. Bull.* **86**, 638–641 (1979).

5. Markwick, B. *Proc. Soc. Psychical Res.* **56**, 250–280 (1978).

Tough time for taxonomy

Sir—In support of Henry Disney's letter¹ highlighting the importance of taxonomy, may I add the following. In the field of bioprospecting², whether for natural products or genes, the skills of a field taxonomist are indispensable. This is true

for a number of reasons, the most important being the issue of re-collection.

If an organism has yielded a natural product with interesting properties, in many cases the species needs to be identified and re-collected. Without an alpha taxonomist present, it is highly likely that the wrong species may be re-collected. In addition, the skills of the field taxonomist are invaluable in order to collect related species, and to avoid accidental re-collection of species of low value. Bioprospecting relies heavily on alpha taxonomy, as taxonomic identification using molecular biology methods is not an option during a collection expedition.

In my field, marine natural product chemistry³, there are very few marine invertebrate taxonomists left, and it is conceivable that without the necessary investment we will soon be left without anyone capable of identifying species based on morphology.

Once it is realized how important the description of species is in the hunt for novel pharmaceuticals and genes, perhaps industry and the academic community will once again focus on traditional taxonomy, which should complement, and not be replaced by, molecular biology methods. Let us hope that we can achieve a renewal of interest in traditional taxonomy before it is too late.

Marcel Jaspars

Department of Chemistry, Marine Natural Products Laboratory, University of Aberdeen, Meston Walk, Old Aberdeen AB24 3UE, UK

1. Disney, R. H. *Nature* **394**, 120 (1998).

2. Macilwain, C. *Nature* **392**, 535–540 (1998).

3. Jaspars, M. in *Advances in Drug Discovery Techniques* (ed. Harvey, A. L.) (Wiley, New York, 1998).

How to simplify the plutonium problem

The global stockpile of separated plutonium is expensive and hazardous to reprocess, vulnerable to terrorist threat and disposal is costly. But there are apparently simple ways to reduce this problem.

Frank N. von Hippel

As a result of the reduction of their nuclear arsenals following the end of the Cold War, the United States and Russia have each declared almost half of the plutonium they produced for weapons. The United States has started a \$2 billion programme to dispose of 50 tonnes of this excess plutonium, either by using it in fuel for power reactors or by mixing it with high-level radioactive waste. Both approaches embed the plutonium in a massive matrix surrounded by a long-lived (30-year half-life) lethal radiation field, making it much less vulnerable to theft.

Russia has also committed itself to disposing of up to 50 tonnes of weapons plutonium by turning it into reactor fuel, if given sufficient financial and technical assistance by the G-7 group of Western countries.

Meanwhile, France and Britain are operating multi-billion-dollar reprocessing programmes to separate about 20 tonnes of plutonium a year from the spent fuel of civilian nuclear power reactors. About 9 tonnes were recycled into fuel in 1997. But the accumulated global stockpile of separated civilian

plutonium still grew to about 170 tonnes — comparable to the superpower stockpiles of military plutonium¹.

So two enormously costly activities, those of embedding and of separating plutonium from a protective matrix for recycling, are proceeding in parallel. Plutonium separated from fuel in nuclear power reactors can easily be stolen and is directly usable in weapons. Although it will be difficult to change the current situation, the need to do so urgently is obvious, in view of the huge waste of time and resources, as well as the handling risks and the threat of terrorism. I make proposals here for policy changes to reduce these risks.

Myth of plutonium differences

The spent-fuel reprocessing industry attempts to justify its plutonium-separation activities by invoking the difference in isotope make-up between 'weapons-grade' and 'power-reactor-grade' plutonium. For example, Jean-Pierre Rougeau of Cogema, the French reprocessing company that accounts for more than half of the world's plutonium-separation activity, said at a conference in 1994 that "reactor-grade plutonium is not realistically a potential weapons material... it is — practically speaking — virtually impossible to convert reactor-grade plutonium to military use". The following year I was told by a senior official at France's Commission on Atomic Energy that its division of military applications totally disagreed with Cogema's statement.

According to US weapons experts, "a potential proliferating state or subnational group using designs and technologies no more sophisticated than those used in first-generation nuclear weapons could build a nuclear weapon from reactor-grade plutonium that would have an assured, reliable

yield of one or a few kilotonnes [1,000 tonnes chemical explosive equivalent].... Advanced nuclear weapon states such as the United States and Russia, using modern designs, could produce weapons from reactor-grade plutonium having reliable explosive yields, weight, and other characteristics generally comparable to those of weapons made from weapons-grade plutonium"².

Misinformation on this question even crept into the UK Royal Society's report of February 1998 on the problem of Britain's plutonium stockpile — now 50 tonnes but projected to grow to 100 tonnes by 2010. The society's study group apparently compared the weapons usability of reactor-grade plutonium oxide, the chemical form of plutonium stored at Britain's commercial reprocessing plant, with that of weapons-grade plutonium metal, the form used in weapons. This comparison is irrelevant, as it is easy to convert plutonium oxide to metal.

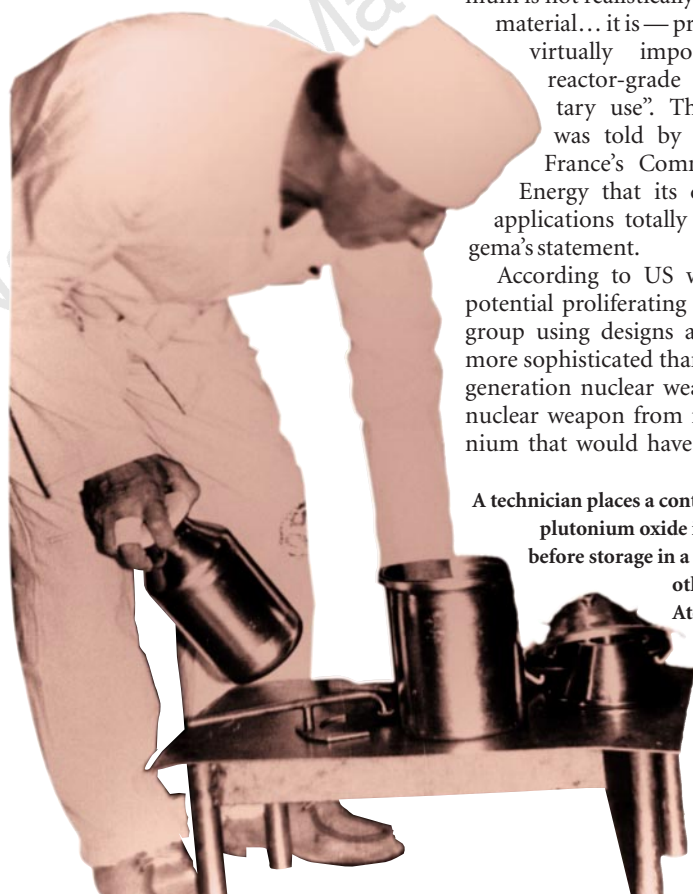
Despite this error, the report concluded that reactor-grade plutonium is "a plausible target for determined terrorist groups or states wanting to make nuclear weapons". It requested the British government to commission a comprehensive review by independent experts of options for disposing of some of the plutonium in the stockpile and to take steps "to reduce the amount being added to it each year, primarily by reducing the amount of reprocessing".

Demand for reprocessing

Commercial reprocessing in France and Britain has been financed mainly by other countries, particularly Japan and Germany. Why do these countries want costly reprocessing services when the recycling of the separated plutonium into mixed-oxide (MOX, plutonium-uranium) fuel must be subsidized and the relicensing of reactors to use MOX fuel is often prolonged and contentious?

The original justification for civilian reprocessing was to provide start-up plutonium for a planned new generation of plutonium 'breeder' reactors that would multiply by 100 times the energy recoverable from natural uranium by converting abundant non-chain-reacting ²³⁸U into chain-reacting plutonium. But the enormous growth projected for nuclear power did not materialize, and enough high-grade uranium ore has been found to eliminate the need for costly breeder reactors for the foreseeable future.

During the 1960s and 1970s, however, the vision of a plutonium-based energy economy was embraced by virtually all the



A technician places a container holding 2.8 kg of plutonium oxide into an outer container before storage in a warehouse with 12,000 others. The International Atomic Energy Authority considers 8 kg of plutonium sufficient for a first-generation nuclear weapon. (Photo taken by the author during a visit by US representatives in October 1994 to the Mayak reprocessing plant in Russia.)

nuclear establishments of the world — and used by many of them as a convenient cover for the development of nuclear weapons. It was by this route, for example, that US-trained Indian engineers separated the plutonium that was used in India's 'peaceful' nuclear explosion of 1974 and for five nuclear-weapons tests in May this year.

Argentina, Brazil, France, Germany, Iraq, South Korea, North Korea, Pakistan, Sweden and Taiwan have all pursued the development of nuclear weapons under the cover of 'civilian' plutonium programmes. Fortunately, internal political changes and external pressures have aborted most of these programmes, but the danger is not past and indeed is exacerbated by the commercial interests of the reprocessors. Britain's government-owned reprocessing company, British Nuclear Fuels Limited (BNFL), and Russia's Ministry of Atomic Energy, for example, continue to encourage South Korean interest in reprocessing in the hopes of receiving contracts.

Commercial reprocessing continues principally because it has been politically difficult for the nuclear utilities of Japan, Germany and other countries to find domestic locations for long-term storage of spent reactor fuel. Reprocessing contracts with France and Britain have allowed these utilities to export the problem for a decade or more.

National policy has blocked US utilities from following this example, so they have been forced to construct additional storage capacity at their nuclear power plants. Other countries have begun to do the same. Local groups, of course, resist expanded on-site storage out of concern that the sites could turn into long-term radioactive-waste storage sites after the nuclear power plants have been shut down.

Dangers of plutonium 'diversion'

The separation of plutonium for whatever purpose creates the risk of theft by terrorist groups, however stable the host country. Since the collapse of the Soviet Union, however, concern about the danger of plutonium 'diversion' has focused principally on Russia. The United States has mounted large efforts to help Russia upgrade its materials protection, accounting and control systems and is co-funding the construction of a high-security storage facility for excess Russian weapons plutonium.

Almost ignored is the fact that Russia is continuing to operate the Mayak commercial reprocessing plant in the Urals, separating 1–2 tonnes of plutonium a year from Hungarian, Russian and Ukrainian spent reactor fuel, and adding the recovered plutonium to a 30-tonne stockpile in a 50-year-old warehouse (pictured on previous page). The G-7 countries are not encouraging Russia to suspend this reprocessing operation — perhaps

because that would bring into question similar activities in Britain, France and Japan. And, after being criticized by government officials from these countries for the past 20 years, officials of the US State Department are afraid to raise the topic of reprocessing.

How to solve the problem

The enormous accumulation of separated civilian plutonium represents a global security threat that can no longer be ignored. The surest anti-proliferation measure is to stop reprocessing spent fuel and to reduce the quantity of separated plutonium in store. Given the billions of dollars of reprocessing contracts held by British, French and Russian reprocessing companies, how is this to be done?

For Britain, there is, in theory at least, a simple and elegant way out. In 1993, while I was the assistant director for national security in the White House Office of Science and Technology Policy, I tried without success to persuade the State Department to suggest it to the British government.

At the time, BNFL was preparing to put into operation its new commercial thermal oxide reprocessing plant (THORP) at Sellafield, in the north of England. THORP had been financed largely with prepaid reprocessing contracts with foreign utilities — about 60 per cent from Japan and the rest from Western Europe. The essence of these contracts was that Britain would take 4,500 tonnes of foreign spent light-water-reactor fuel and some years later return to the countries of origin the 40 tonnes of contained plutonium and containers of glass containing the equivalent of the radioactive waste produced in the process³.

My proposal, therefore, was that, without operating the new reprocessing plant, Britain should exchange 40 tonnes of its separated civilian plutonium and the associated radioactive waste for the more easily stored foreign spent fuel. Britain would save not only the operating costs of the THORP plant but its multi-billion-dollar end-of-life decontamination cost as well.

Today, the savings in decontamination costs are no longer available. But Britain could still trade perhaps 30 tonnes of its stockpile of separated reactor-grade plutonium and the associated radioactive waste for the remaining unprocessed foreign spent fuel. Because virtually all the foreign fuel is already in storage in Britain, no additional spent fuel-storage capacity would be required.

Britain could also suspend the reprocessing of the fuel from its own advanced gas-cooled reactors. This fuel is as suitable as light-water-reactor fuel for long-term storage. The cost of building spent-fuel storage is considerably less than that of reprocessing, and the costs of storing additional plutonium and radioactive waste would be

avoided. This would reduce Britain's stockpile of reactor-grade plutonium by an additional 25 tonnes by 2010.

With these two initiatives, Britain would reduce its plutonium stockpile problem by more than half by 2010. It could potentially follow the US example and mix the rest of the stockpile with its backlog of liquid high-level waste as that waste is glassified.

France does not have a large domestic stockpile of plutonium with which it could carry out similar trades, but it could offer interim storage of foreign spent fuel as an alternative to reprocessing services. Cogema already has contracts with German utilities that leave open for some years the option of shipping unprocessed spent fuel back to Germany after interim storage in France.

Russian reprocessing contracts currently leave the ownership of separated plutonium in Russia. Although it is against Russian environmental law, Mayak has been tacitly agreeing to keep the radioactive waste from reprocessing as well. Its gross earnings from its relatively small-scale reprocessing activities are only tens of millions of dollars a year, however, and it might be willing to suspend them in exchange for G-7 financing of jobs to store securely and then dispose of Russia's excess plutonium. Of course, changing existing reprocessing contracts would not be easy. Both sides would have to be willing to do so.

From the point of view of the reprocessing countries, the issue would be the loss of jobs, whereas for Britain's foreign customers, a trade of their spent fuel for British plutonium and reprocessing waste might be the equivalent of what they contracted for — but, if the reprocessing contracts are renegotiated, they are likely to prefer the return of their spent fuel after a decade or two of interim storage abroad. Disposing of the separated plutonium is costly, and storage of reprocessing waste is as troublesome as storage of unprocessed spent fuel. These same considerations make it unlikely that the foreign demand for reprocessing services will persist for much longer.

In the end, Britain will probably have to dispose of its huge stockpile of separated plutonium at home. However, if it waits until reprocessing has ended and all the reprocessing waste that could serve as a protective matrix for the plutonium has been glassified, it will have to resort to much more costly disposal options. □

Frank N. von Hippel is at Princeton University, Princeton, New Jersey 08544, USA.
e-mail: fvhippel@princeton.edu

1. Noburu, O. *IAEA Bulletin* 40, 12–17 (1998).
2. US Department of Energy *Nonproliferation and Arms Control Assessment of Weapons-Usable Fissile Material Storage and Excess Plutonium Disposition Alternatives* 38–39 (US Dept of Energy, 1997).
3. Albright, D. et al. *Plutonium and Highly Enriched Uranium 1996: World Inventories, Capabilities and Policies* 162 (Oxford Univ. Press, Oxford, 1997).

A sense for landmines

Anthony W. Czarnik

A heightened sense of smell, achieved by proxy through chemistry, offers solutions to problems where biologically based sensors are inappropriate. Detection of the TNT in landmines using fluorescent chemosensors is one example.

If the human sense of smell were as keen as a dog's, our perception of the world would be very different. A child could sense its mother in another room. A border agent would know when drugs were being smuggled in the car passing by. And a soldier clearing a field of landmines could avoid stepping or kneeling on the one buried just in front of him (Fig. 1). Of course, most dogs' sense of smell is much better than ours, and there's no immediate prospect that this biological fact will be undone. But non-biological enhancement of our sense for odours can be achieved with sensors — devices, molecule-sized or larger, that signal the presence of matter (or energy) in a way discernible by one of the human senses.

Writing in the *Journal of the American Chemical Society*¹, Yang and Swager describe the evaluation of porous, shape-persistent, fluorescent polymer films that serve as chemosensors for explosives — in particular, trinitrotoluene (TNT) and its derivatives, which are commonly used in landmines. Sensors for TNT are in routine use at airports, for instance, but the technology requires sophisticated instruments and simply is not rugged enough for field use. And other ways of detecting landmines are far from ideal — dogs tire easily and, with their requirement for a full-time trainer, are expensive; ground-penetrating radar and metal detectors lack the specificity of chemical sensors and produce too many false positives.

Chemosensors — sensors incorporating a receptor element that is not derived from biology^{2,3} — have enjoyed increasing attention as alternatives to biosensors. Occasionally, the alternative is required because no biological receptor has been discovered for a given analyte (the chemical species one wishes to detect). More often, the alternative is needed because the biological receptor does not have the properties that are ultimately useful. Although biological macromolecules such as antibodies have deliciously high affinity for many analytes, they are not very rugged molecules. Shake most proteins too hard, and they lose activity. Heat most proteins to too high a temperature, and they denature. Eat most proteins, and they are hydrolysed to amino acids.

For the soldier in the field running mine-



Figure 1 Bosnia, 1995 — United Nations troops sift through snow, ice and mud in a search for landmines.

clearance programmes, such delicacy is a real problem. Refrigeration is rarely available, and equipment inevitably gets bashed about. In the case of landmine detection, the chance of a false negative stemming from 'dead' antibody in the sensor presents an intolerable risk. So although antibodies to TNT exist, this biological receptor has not yet been used in a field device. Instrumental detection methods, such as mass spectrometry, are ultra-sensitive, and there have been impressive advances in portability. But they suffer from the same lack of in-the-field ruggedness.

If it is ultimately impossible to make antibodies that are robust enough, then the search for receptors must begin with ruggedness. Synthetic receptors, which need not rely on a metastable three-dimensional structure or hydrolysable backbone, fit the bill. A pH indicator strip incorporates a 'receptor' for protons; that receptor is very rugged, and bottled strips can sit on the shelf for years yet retain their activity. The study of synthetic receptors is not new; indeed, there is an extensive literature of non-applied research on the topic. The literature on work that has been 'reduced to

practice', by contrast, is much more limited (examples include the stabilization of pharmaceutical products through the inclusion of cyclodextrin and the sensing of calcium-ion concentrations in living cells).

In the sensor business, 'reduced to practice' means, 'meets all the demands of the application' — no more, no less. For TNT sensing above buried landmines, those demands are predictable:

- Sensitivity (that is, transmission of a stable signal resulting from the sensing of concentrations of airborne TNT of just a few parts per billion).
- Selectivity (discrimination of those low concentrations of TNT from much higher concentrations of other airborne chemical species).
- Real-time signal development (the signal appears within a few seconds, even in the cold; faster is better).
- Range (the sensor does not become saturated at lower than the usual concentration of analyte).
- Reversibility (remove the TNT and the sensor returns to its initial state within a few seconds; again, faster is better).
- Stability (tolerant of water; but active for months in the solid state at high temperatures, such as those typical of deserts).

Yang and Swager's work¹ has not yet been 'reduced to practice', but it is a necessary and innovative step towards that goal. Swager's laboratory previously showed that conjugated, polymer-based fluorescent chemosensors can provide signal amplification for the innately sensitive fluorimetric method⁴, which involves the synthesis of luminescent polymers that are quenched by analyte molecules along the entire length of the polymer. However, such films of flat molecules suffered from 'pi-stacking' and the fluorescence self-quenching that accompanies it. Yang and Swager have now made films of polymers derived from fluorescent pentaerythritol molecules that are not flat. As predicted, this film showed less quenching than the first film (although the two systems are not perfectly comparable). But quenching does occur when the flat TNT molecule is introduced. The authors postulate that the three-dimensional pentaerythritol polymer has pores, into which the flat TNT molecule can enter and quench fluorescence. This quenching results in a decrease in fluorescence intensity, which serves as the physical basis for signal transduction in the presence of TNT (Fig. 2, overleaf). Several other flat organic molecules, such as benzophenone and 1,4-dicyanobenzene, did not induce fluorescence quenching when tested — that is, the sensor film can discriminate between TNT and other molecules of similar size and electronic character.

To what extent does Yang and Swager's scheme meet the demands necessary for a field TNT sensor? The reported signal amplitude of $\times 20$, real-time signal development

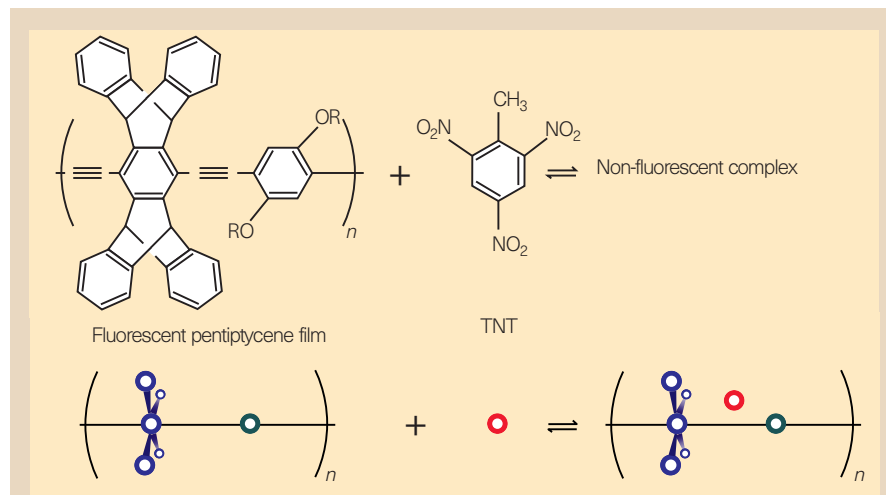


Figure 2 Chemistry of the prototype TNT chemosensor devised by Yang and Swager¹ (top), with the geometrical relationships being depicted below. The scheme uses a fluorescent polymer film, and because the pentiptycene molecules are not all flat, the film contains pores into which TNT molecules can enter. Binding of a TNT molecule to the polymer results in quenching of fluorescence — the non-fluorescent complex shown here. The consequent change in intensity can be easily measured, and serves as a signal indicating the presence of TNT. As discussed in the text, this approach has yet to be 'reduced to practice'.

and reversibility (through rinsing) are all encouraging. The sensitivity to TNT is not yet described, however. And although the discrimination against several planar aromatic compounds is good news, field use may reveal interference by other, as-yet-unknown, compounds; one emerging approach to such challenges of selectivity is the sensor-array format^{5–7}, by which large collections of nonspecific chemosensors all interact with a sample, and the analyte concentration is extracted by network analysis.

None of these drawbacks is cause for pessimism. Research on chemosensors is just starting. Each new discovery adds to the toolbox that will help all of us smell better —

and in this case we can hope, in due course, for a viable field detector for TNT. □

Anthony W. Czarnik is at Illumina Inc., 2187 Newcastle Avenue, STE 101, Cardiff, California 92007, USA.

e-mail: aczarnik@illumina.com

1. Yang, J.-S. & Swager, T. M. *J. Am. Chem. Soc.* **120**, 5321–5322 (1998).
2. Yoon, J.-Y. & Czarnik, A. W. *J. Am. Chem. Soc.* **114**, 5874–5875 (1992).
3. Czarnik, A. W. *Accs Chem. Res.* **27**, 302–308 (1994).
4. Zhou, Q. & Swager, T. M. *J. Am. Chem. Soc.* **117**, 7017–7018; 12593–12602 (1995).
5. Michael, K. L., Taylor, L. C., Schultz, S. L. & Walt, D. R. *Analyt. Chem.* **70**, 1242–1248 (1998).
6. Lavigne, J. J. *et al. J. Am. Chem. Soc.* **120**, 6429–6430 (1998).
7. Doleman, B. J., Sanner, R. D., Severin, E. J., Grubbs, R. H. & Lewis, N. S. *Analyt. Chem.* **70**, 2560–2564 (1998).

the cause of population declines as they occurred rather than after the event. They found that the same pathogen, a chytrid fungus, is the cause of mortality at the two localities. The fungus invades the skin of adult frogs and probably causes death by interfering with their cutaneous respiration and water uptake. It particularly infects the pelvic patch, an important site of water absorption in many frogs.

There are a number of remarkable features about this discovery. First, the chytrid fungus is a newly discovered genus. Rather than having a typical branched, filamentous structure, it resembles a protozoan and was originally misidentified as such. Second, although chytrids are widespread in soil and are known to attack plants and insects, this is the first to be identified as a vertebrate pathogen. Third, this fungus has now been found in two distant natural localities, as well as among captive amphibians in zoos and aquaria in Australia and the United States. Finally, the chytrid seems to attack just adult amphibians (Fig. 1), having no harmful effects on their tadpoles. This is probably because it only attacks skin that contains the protein keratin, which does not occur in amphibians until metamorphosis. This feature differentiates Berger and colleagues' study from many previous reports of amphibian population declines, a characteristic of which has been reproductive failure owing to mortality among eggs and larvae.

Although this is a very exciting discovery, claims in the popular media⁷ that the answer to the declining amphibian conundrum has been found are premature. It does not solve the puzzle, but it does raise further questions. For example, is the chytrid fungus ubiquitous, or has it only recently found its way to Australia and Panama? If it is widespread, why have amphibians only just become susceptible to it? If it has been introduced, where has it come from and how is it being spread? Until we can answer these questions, we will not be able to assess the significance of the new findings in the context of amphibian declines. Of immediate concern is the alarming possibility that herpetologists, seeking evidence for causes of declines in natural populations, may be unwittingly helping to spread disease on footwear or collecting gear⁷.

This is not the first time that disease has been implicated in amphibian population declines. The ubiquitous bacterium *Aeromonas hydrophila*, which causes the condition known as 'red leg', has been a contributory factor in some cases^{8,9}, and iridoviruses have been implicated in mass frog mortalities in the United Kingdom¹⁰. All of these events may be linked not by the specific pathogen, but by the possibility that the immune systems of many amphibian species are being compromised by environmental factors such as climate change, chemical contamination or increased levels of ultra-

affected sites some species are declining whereas others are not. Finally, there is no single cause for these declines.

This last point is now addressed by Berger *et al.*⁴, reporting in *Proceedings of the National Academy of Sciences*. They have identified a pathogen that links declines among frog populations in two such geographically distinct parts of the world — Panama and Queensland, Australia. Population declines have been well documented at both of these localities^{5,6} and, in an exemplary example of international scientific collaboration, biologists from Australia, the United Kingdom and the United States have come together to examine the material collected from the two areas.

Berger and colleagues studied many dead and dying frogs; this allowed them to identify

Ecology

A declining amphibian conundrum

Tim Halliday

Over the past decade there have been dramatic declines among some amphibian populations in many parts of the world, including a number of apparent species extinctions¹. These events have caused particular concern because many have occurred in protected areas such as nature reserves and national parks^{2,3}. Moreover, until recently they have been detected only after they have happened. The Declining Amphibian Populations Task Force — a worldwide network of scientists, of which I am an international director — has been investigating the declines since 1991, and we have reached several conclusions. First, amphibian declines are occurring throughout the world, although some regions are not affected. Second, at most



Figure 1 Decline and fall — this tree frog (*Litoria caerulea*) is one of the amphibian species that Berger *et al.*⁴ have found to be infected with the chytrid fungus.

violet radiation⁸. Amphibian declines seem to have many causes, and it is vital that scientists do not relax their attempts to investigate these other factors — for which there is mounting evidence — in the mistaken belief that this new study has solved the puzzle.

The significance of Berger and colleagues' discovery⁴ may go beyond the mystery of amphibian declines — this is the first time that an infectious disease pandemic has been implicated in the decline and possible extinction of animals. Moreover, if the chytrid pathogen was accidentally introduced into previously naive populations, as measles was to South America, we may be seeing a new kind of anthropogenic insult to the environment. This threat may be more

insidious, and difficult to control, than existing environmental problems such as chemical pollution or habitat destruction. □

Tim Halliday is in the Department of Biology, The Open University, Milton Keynes MK7 6AA, UK.
e-mail: t.r.halliday@open.ac.uk

1. Wake, D. B. *Science* **253**, 860 (1991).
2. Pounds, J. A., Fogden, M. P., Savage, J. M. & Gorman, G. C. *Conserv. Biol.* **11**, 1307–1322 (1997).
3. Drost, C. A. & Fellers, G. M. *Conserv. Biol.* **10**, 414–425 (1996).
4. Berger, L. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 9031–9036 (1998).
5. Lips, K. R. *Conserv. Biol.* **12**, 106–117 (1998).
6. Richards, S. R., McDonald, K. R. & Alford, R. A. *Pacific Conserv. Biol.* **1**, 66–77 (1993).
7. Anderson, I. *New Sci.* No. 2140, 4–5 (1998).
8. Carey, C. *Conserv. Biol.* **7**, 355–362 (1993).
9. Bradford, D. E. *J. Herpetol.* **25**, 174–177 (1991).
10. Cunningham, A. A. *et al.* *Phil. Trans. R. Soc. Lond. B* **351**, 1539–1557 (1996).

Superconductivity

Strain yourself

Ivan K. Schuller

On page 453 of this issue¹, J.-P. Locquet and colleagues show that the superconducting transition, or critical, temperature (T_c) of a particular copper oxide can be doubled by the clever use of epitaxial strain. Increases in transition temperature using pressure have been accomplished before, but the increase obtained here is much larger than any achieved by using standard pressure techniques.

Thin solid films have been used for many years to study materials under physical conditions that cannot be achieved in the bulk. Very large pressure can be applied, and the intrinsically different behaviour of lower-dimensional systems can be observed, for example. Films have shown us many new phenomena, including giant magnetoresistance, dimensional crossover in superconducting multilayers, competition between superconductivity and magnetism, and the

high reflectivity of X-ray mirrors^{2–4}. The properties of thin films are different from those in the bulk largely because films are grown artificially, under conditions far from equilibrium (usually using a variety of sophisticated, high-vacuum techniques).

But quantitative structural and chemical characterization of a thin film is difficult, and many studies fail because defects such as roughness or interdiffusion of atoms invalidate the conclusions obtained. Locquet and collaborators have avoided these problems in a systematic and thorough fashion. Their films produce clear, sharp peaks in X-ray diffraction, and atomic-resolution electron microscopy shows a low defect density. (Generally, defects tend to depress the superconducting transition temperature, especially in the high- T_c oxides.)

Locquet and colleagues have grown films of $\text{La}_{1.9}\text{Sr}_{0.1}\text{CuO}_4$ on substrates carefully

chosen to be of slightly smaller lattice constant (atomic spacing) than the natural value for $\text{La}_{1.9}\text{Sr}_{0.1}\text{CuO}_4$. Because the layers are very thin, the $\text{La}_{1.9}\text{Sr}_{0.1}\text{CuO}_4$ adopts this lattice constant, putting itself under considerable compressive strain to do so. It is not surprising that strain should affect T_c , because pressure has a measurable effect on ceramic superconductors. But the size of the increase in $\text{La}_{1.9}\text{Sr}_{0.1}\text{CuO}_4$ is striking. The authors measure a factor-of-two change in the superconducting T_c of their films, compared with that of the bulk alloy — an increase from 25 to 49 K.

The strain-dependence of T_c has opposite signs along different crystallographic orientations, so applying hydrostatic pressure (which is necessarily isotropic) may not increase T_c , and could even decrease it. But by an appropriate choice of the substrate material and growth temperature, Locquet and co-workers could apply uniaxial strain only in the directions that increase T_c .

In addition, this work may help to determine the critical structural parameters that control superconductivity, and perhaps even to clarify the mechanism that gives rise to superconductivity. Unfortunately, the results of these experiments alone cannot make a definitive statement in this direction. But they do have one intriguing and controversial implication: if the compression in the plane of the film gives rise to an expansion perpendicular to it (the 'Poisson effect') (Fig. 1), that would imply that the superconducting T_c increases with increasing distance between CuO_2 planes. This is contrary to several experimental and theoretical claims^{5,6}. But some caution should be exercised with this conclusion, as the Poisson effect does not always occur in thin films. Furthermore, experiments generally find that separating two CuO_2 planes from one another decreases

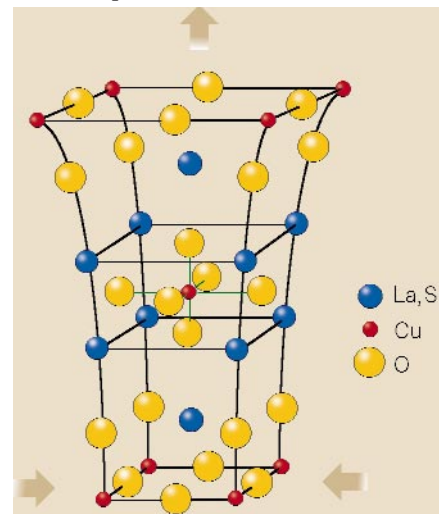


Figure 1 $\text{La}_{1.9}\text{Sr}_{0.1}\text{CuO}_4$ on a substrate that forces it to distort its usual crystal structure (the actual strain is exaggerated here). This distortion can increase the substance's superconducting transition temperature by a factor of two.

the transition temperature in the case of yttrium–barium–copper oxides, and T_c is claimed to be independent of distance⁵ in BiSCCO, a particular bismuth compound. So the question of what transition temperature a single CuO₂ plane would have is still unsolved.

The present results¹ clearly demonstrate that the application of large uniaxial strain can increase superconducting transition temperature by a substantial factor, and perhaps this technique can be extended to other superconductors and lead to even higher transition temperatures.

Practical devices based on thin films are always grown on properly chosen substrates.

Thus a judicious choice of substrates could considerably benefit applications using films that are thin enough to be subject to large epitaxial strains. □

Ivan K. Schuller is in the Physics Department, University of California at San Diego, La Jolla, California 92093-0319, USA.
e-mail: ischuller@ucsd.edu

1. Locquet, J.-P. *et al.* *Nature* **394**, 453–456 (1998).
2. Falicov, L. M. *et al.* *J. Mater. Res.* **5**, 1299 (1990).
3. Dhez, P. & Weisbuch, C. (eds) *Physics, Fabrication and Applications of Multilayered Structures* (Plenum, New York, 1988).
4. Shinjo, T. & Ono, T. *J. Mag. Mag. Mater.* **177**, 31 (1998).
5. Choy, J. H., Kwon, S.-J. & Park, G.-S. *Science* **280**, 1589–1592 (1998).
6. Chakravarty, S. *et al.* *Science* **261**, 337–340 (1993).

Immunology

The original sin of killer T cells

Andrew J. McMichael

The phrase ‘original antigenic sin’ was first used to describe the antibody response to influenza virus. After an initial infection, reinfection (or vaccination) with a new strain of the virus boosted the concentration of antibodies specific for the earlier infecting strain^{1,2}. Although these antibodies cross-reacted with the new virus, they had higher affinity for the original infecting strain. This was immediately seen to have big implications for vaccine design — a vaccine based on a new strain of influenza virus might be unable to prime antibodies to the intended virus in people who had already been infected with a related viral strain.

Until now, original antigenic sin has been regarded as largely an antibody phenomenon. But, on page 482 of this issue, Klenerman and Zinkernagel³ describe original sin in the response of cytotoxic T lymphocytes (CTLs) to lymphocytic choriomeningitis virus (LCMV). Their work was stimulated by studies of people with the human immunodeficiency virus (HIV), who sometimes mount a CTL response to an immunodominant strain of the virus with weak or no response to the other immunogenic variants present^{4,5}.

The authors attacked the problem *in vivo*. They infected mice with strains of LCMV that were either normal (wild type) or mutated at the immunodominant epitopes recognized by the CTL. For each mutated strain studied, they found that the CTL response was asymmetrical. Mice infected with the wild-type virus showed only a weak cross-reactive specificity when challenged with the mutant strain, reacting mainly to the wild-type virus. The low reactivity of CTLs against the mutant strain was also associated with delayed clearance of this

strain by the immune system. In contrast, the CTL responses from mice infected with the mutant virus cross-reacted equally with the mutant and wild-type strains. Thus, the order in which the mice are exposed to different variants of the virus could have a significant effect on the outcome of the infection.

How does this happen? A possible explanation⁶ for antibody original antigenic sin emerged when helper T cells were discovered, and it was realized that T cells react with peptide fragments of viral proteins that are often conserved between different strains. So, helper T cells, primed by the original virus and then stimulated by the new infec-

tion, might activate memory B cells that are specific for the original virus (although it is not clear why antibodies specific for the new variant are not stimulated too). In the same way, the phenomenon observed by Klenerman and Zinkernagel³ might reflect the very strong CTL memory response to LCMV infection. Studies^{7–9} of acute LCMV infection in mice showed that massive CTL responses occurred, whereby 25–50% of all CD8-positive T cells were virus specific. Even after recovery, 10% of peripheral CD8-positive T cells were specific for the immunodominant LCMV epitopes. A weakly cross-reacting new virus might, therefore, reactivate these plentiful CTL memory cells more readily than the much less abundant naive CTL precursors (Fig. 1). Although this mechanism depends on very high levels of memory CTLs, similar numbers are seen in persistent HIV infection¹⁰, where original antigenic sin may also occur^{4,5}.

A second possibility is that activated memory CTLs remove enough of the antigen-presenting cells (APCs) to abort the primary CTL response to the new viral epitope. This would be consistent with the delayed clearance of mutant compared with wild-type virus in secondary challenges. The memory CTL could simply kill the APC, although recent results^{11–13} indicate another possibility. A type of APC called the dendritic cell must be activated in order to initiate CTL responses. Activation occurs through an interaction between the CD40 ligand on T helper cells and CD40 on the dendritic cells. But if this requirement for activated dendritic cells is less stringent for the memory CTLs, these CTLs could deactivate the dendritic cells, thereby inhibiting new

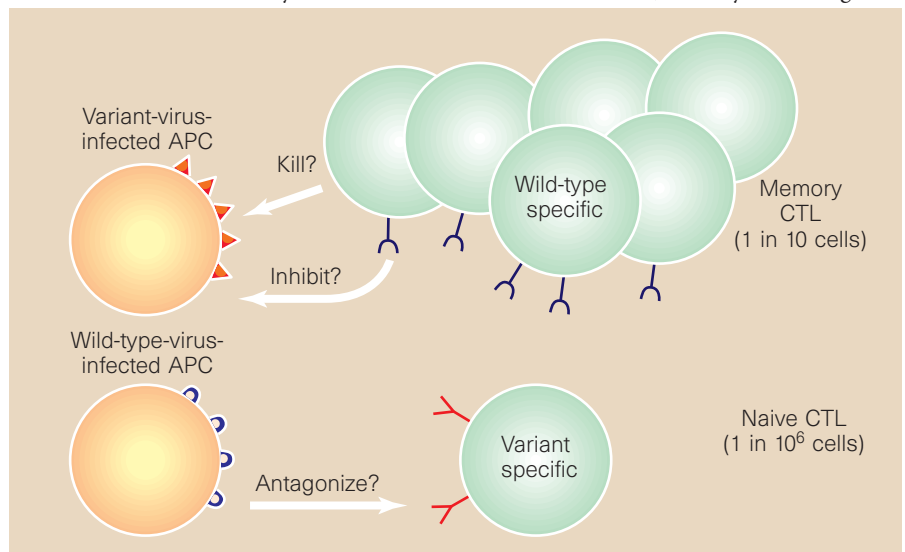


Figure 1 Original antigenic sin in cytotoxic T lymphocytes (CTLs). Klenerman and Zinkernagel³ have discovered that the phenomenon of original antigenic sin — whereby infection with a virus boosts the concentration of antibodies against a related strain from a previous infection — also occurs in the response of CTLs. Memory CTLs stimulated by wild-type antigen (blue) are abundant and, although weakly stimulated through their receptors, can kill or inhibit antigen-presenting cells (APCs) presenting the variant virus (red). Persisting APCs that present the wild-type virus may partially activate and antagonize naive T cells specific for the variant virus.

primary CTL responses. Deactivation could occur through the interferon- γ and other cytokines that are released from memory CTLs within a few hours of contact with antigen¹⁴.

Another explanation would be that a form of T-cell-receptor antagonism occurs¹⁵. If the CTL clones that react with the mutant viral epitope have receptors that bind sub-optimally to the wild-type epitope, the result might be partial activation, leading to anergy (a failure of the immune system to respond). Thus, the wild-type virus would antagonize the primary CTL response to the mutant strain. For this mechanism to work, antigen from a previous infection would need to persist. Low levels of integrated DNA persist in mice infected with LCMV¹⁶, although these levels are probably too low for antagonism to occur. Moreover, in Klenerman and Zinkernagel's study, the CTLs stimulated by the mutant virus cross-reacted well with the wild-type virus, arguing against this explanation.

Original antigenic sin in CTL responses could be very important — it could explain some of the complexity in CTL responses to HIV and its variants^{17,18}. If it were a common phenomenon, it would give viruses another means of escape from the immune system. That is, once a high-level CTL response has become fixed, immunogenic mutants might escape and become dominant in the swarm of viral quasispecies, failing to provoke an effective CTL response.

Just as in the initial description of original antigenic sin, the new work³ has implications for vaccine design. Many groups are now trying to design CTL-inducing vaccines for the

control of variable viruses such as HIV and hepatitis C virus. Original antigenic sin means that a monovalent vaccine (such as a peptide) intended to stimulate CTLs may not work if the virus varies at that epitope. Not only would the vaccine-induced CTL response fail to control infection with any variant virus, but the CTL response to the equivalent epitope in that virus might also be impaired. The vaccine might even make the infection worse. The probable solution is to choose several well-conserved epitopes for the vaccine — but HIV is so variable, these may be in short supply. □

Andrew J. McMichael is at the Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK.

e-mail: amcmichael@hammer.imm.ox.ac.uk

1. Francis, T. *Ann. Int. Med.* **39**, 203–221 (1953).
2. Fazekas de St Groth, S. & Webster, R. G. *J. Exp. Med.* **124**, 331–346 (1966).
3. Klenerman, P. & Zinkernagel, R. M. *Nature* **394**, 482–485 (1998).
4. McAdam, S. N. *et al. J. Immunol.* **155**, 2729–2736 (1995).
5. Klenerman, P., Meier, U. C., Phillips, R. E. & McMichael, A. J. *Eur. J. Immunol.* **25**, 1927–1931 (1995).
6. Askinas, B. A., McMichael, A. J. & Webster, R. G. in *Basic and Applied Influenza Research* (ed. Beare, A. S.) 157–188 (CRC Press, Boca Raton, FL, 1982).
7. Butz, E. & Bevan, M. *Immunity* **8**, 167–175 (1998).
8. Gallimore, A. *et al. J. Exp. Med.* **187**, 1383–1393 (1998).
9. Murali-Krishna, K. *et al. Immunity* **8**, 177–187 (1998).
10. Ogg, G. S. *et al. Science* **279**, 2103–2106 (1998).
11. Bennett, S. R. M. *et al. Nature* **393**, 478–480 (1998).
12. Ridge, J. P., DiRosa, F. & Matzinger, P. *Nature* **393**, 474–478 (1998).
13. Schoenberger, S. P., Toes, R. E. M., van der Voort, E. I. H. & Melief, C. J. M. *Nature* **393**, 480–483 (1998).
14. Lalvani, A. *et al. J. Exp. Med.* **186**, 859–865 (1997).
15. Meier, U.-C. *et al. Science* **270**, 1360–1362 (1995).
16. Klenerman, P., Hengartner, H. & Zinkernagel, R. M. *Nature* **390**, 298–301 (1997).
17. Haas, G. *et al. J. Immunol.* **157**, 4212–4221 (1996).
18. Harter, T. *et al. J. Immunol.* **156**, 2616–2623 (1996).

more than just the geologists in the crowd.

Generally, the characteristics of these abrupt climate shifts implicate changes in large-scale ocean circulation as the culprit: the atmosphere adjusts to perturbation too quickly and does not have the 'memory' required to drive millennial-scale variability; the solid Earth (including polar ice), on the other hand, reacts too slowly. There are conceivable exceptions to this generalization. For example, glacial processes could be involved, because some ice sheets are inherently unstable and are prone to surging². Also, the possibility of external influences, such as changes in solar luminosity or tidal energy, cannot be discounted. However, even unstable ice sheets must respond to some initial climate disturbance, and so glacial processes are themselves probably not the primary agent of change. And any connection with external forces is extremely difficult to prove or test.

If the answer lies in the ocean, then what aspect of ocean circulation could suddenly alter the global climate system, after operating normally for more than a thousand years? The most widely recognized hypothesis has been that the formation of North Atlantic Deep Water — a process by which warm surface currents lose their heat to the atmosphere and sink to fill the deep ocean basin — flipped between two stable modes of operation. For climate purposes, these modes are essentially 'on' or 'off'³. The climatic effect of such a change might be significant: deepwater formation around Greenland today brings oceanic heat northward, making much of Europe several degrees warmer than it probably would be otherwise. And, as a water mass, North Atlantic Deep Water acts to redistribute heat and salt globally.

Furthermore, there is ample evidence from deep-sea sediment cores (much of which was presented at the conference) showing that sinking processes in the North Atlantic did in fact undergo millennial-scale oscillations similar to those inferred from climate records such as the Greenland ice cores (for an example, see ref. 4). So this mechanism, popularly known as the ocean's 'conveyor belt', could explain the observations of abrupt climate change from the entire circum-Atlantic region. It also might explain the connection between the severity of the abrupt shifts and the presence of ice sheets, if the continental ice had a varying influence on the density (the tendency for sinking) of North Atlantic surface waters.

The problem with this seemingly tidy package is that, to drive fluctuations in global temperature, either the greenhouse capacity or the reflectivity (albedo) of the planet must be affected. And without invoking complicated feedbacks, it is not clear from general principles that the North Atlantic circulation has the climatic influence to

Palaeoclimatology

The ends of an era

Christopher Charles

If geologists were to construct 'top ten' lists of the advances of the century, the relatively recent discovery that the Earth has switched repeatedly and abruptly between cold and warm climates over the course of ice-age cycles would no doubt feature prominently. Yet, despite the decade of research devoted to characterizing this so-called 'millennial-scale' climate variability, there are still primary questions about the origin of such abrupt changes and, therefore, equal uncertainties in predicting how climate might behave in the future. These questions surfaced last month at an American Geophysical Union conference*, where a few scientists challenged established dogma centred on the North Atlantic by suggesting that the tropical Pacific may have been

involved — a fitting proposal in a year when El Niño took the blame for a variety of human woes.

The conference objective was to discuss the climate processes that may explain the inferences from archives such as ice cores and deep-sea sediment cores — temperature swings of up to 10 °C, spaced several thousand years apart, occurring over much of the Earth's surface but more violently when the Earth was in a partly or fully glaciated state. The disquieting feature of these climate changes is that the switch between cold and warm conditions apparently happened in as little as 20 years¹. Identifying the processes responsible would help assess the likelihood of climate 'surprises' associated with global warming. And obviously, any similar climate shifts occurring in the near future would have profound repercussions for Earth's habitability. So the subject was of interest to

*Chapman Conference on Mechanisms of Millennial-Scale Global Climate Change, Snowbird, Utah, 14–18 June 1998.
http://www.agu.org/meetings/cc98dcall.html

achieve either of those effects on a large scale.

In this respect, there are observations that cannot be accommodated easily in the conveyor-belt scheme of events. How is it, for example, that millennial-scale variability appears, nearly synchronously, in places as far from the North Atlantic as the Santa Barbara basin (off California)⁵ and the Sulu Sea (Indonesia)⁶? Or that the concentration of methane in the atmosphere (presumably a reflection of microbial activity on tropical land surfaces) varied along with other climate variables⁷? Or that Antarctic air temperatures sometimes varied nearly inversely with those in the Arctic⁸? These observations point to the involvement of other ocean processes — which in turn raises the possibility that the North Atlantic oscillations were themselves merely the by-product of another, even stronger source of variability.

This is where the tropical Pacific enters the discussion. Experiments with a model of climate in the tropical Pacific — in fact, one of those used to predict several of the last El Niños — have revealed that simulated warm anomalies in the Pacific, akin to El Niño warm events, can be sustained for several thousand years and then be terminated abruptly (M. Cane and A. Clement, Columbia Univ.). If the model results are at all close to reality, then they might explain much of the geological evidence, because the global atmosphere is so sensitive to the size and location of the Pacific warm pool, the largest heat engine of the planet. The model experiments were designed to examine the rectified response of the tropical ocean to changes in the seasonality of radiation over tens of thousands of years; the millennial-scale behaviour of the system was a surprise and has yet to be explained. But as a further twist, there is the possibility that the tropical ocean circulation might be modulated by some preconditioning from higher latitudes (G. Philander, Princeton Univ.), which brings in still other ocean processes operating in the subtropics.

Whether real or not, the idea of a millennial-scale 'El Niño' must be given due consideration, because variability in the tropical Pacific is, after all, a known source of global temperature change. As Cane put it, if one were to create a regional climate anomaly in hopes of causing a global disturbance, the best place for that would be the tropical Pacific. Furthermore, the idea connects physical processes that we actually experience every few years to the more abstract geological record. This sensible appeal particularly bears on thinking about future climate. Did 'El Niño' work with, or even switch, the conveyor belt to shape the abrupt changes over the ice ages? The answer makes an enormous difference to projections of the course that global warming may take (for example, the question of whether these changes can happen to us), because the

climate sensitivities of the tropical Pacific and the North Atlantic circulation are so different.

Of course, these issues were not resolved in a week-long conference. Even after years of work, the community is just beginning to approach the full significance of the abrupt shifts in climate. Although the Greenland ice-core records did not exist at the time, the poet Richard Wilbur perhaps anticipated them, reflecting on the disparity between human experience and lessons from the past frozen in the fossil record: "These sudden ends of time must give us pause/we fray

into the future rarely wrought/save in the tapestries of afterthought"⁹. □

Christopher Charles is at the Scripps Institution of Oceanography, La Jolla, California 92093-0220, USA.

e-mail: ccharles@ucsd.edu

1. Severinghaus, J. P. *et al. Nature* **391**, 141–146 (1998).
2. MacAyeal, D. *Paleoceanography* **8**, 775–784 (1992).
3. Broecker, W. S. *et al. Nature* **315**, 21–25 (1985).
4. Marchitto, T. M. *et al. Nature* **393**, 557–561 (1998).
5. Behl, R. J. & Kennett, J. P. *Nature* **379**, 243–246 (1996).
6. Linsley, B. K. *Nature* **380**, 234–237 (1996).
7. Stauffer, B. *et al. Nature* **392**, 59–62 (1998).
8. Blunier, T. *et al. Geophys. Res. Lett.* **24**, 2683–2686 (1997).
9. Wilbur, R. "Year's End" in *New and Collected Poems* 302 (Harcourt Brace Jovanovich, San Diego, 1988).

Marine biology

Sex and the single copepod

Rory Howlett

Consider the ocean as a watery discotheque, thronged with tiny organisms — in particular, copepod crustaceans, dominant constituents of the zooplankton, and desperate to find mates. With an estimated one hundred billion trillion potential partners to choose from, it might seem that copepod mating is trouble free, even given the need to find a partner of the same species, opposite sex and ready to reproduce. As described by papers in a special issue of *Philosophical Transactions of the Royal Society*¹, however, that isn't so.

The reasons are outlined in the preface (by G. Boxshall). The world's oceans consist of 1,347 million km³ of open pelagic water-column, and each cubic metre of sea water may contain just a few copepods. These creatures range in size from just 0.5 mm to 10 mm, and they function at low Reynold's numbers — that is, viscous forces dominate their environment, and copepod movement in water is a bit like human movement in treacle. So it might seem that copepods are

completely at the mercy of water movements.

Why should we care? Copepods are among the most numerous multicelled organisms on Earth², and consume vast amounts of phytoplankton; in turn they are eaten by fish and other predators, thereby transferring carbon to higher trophic levels. So copepod population dynamics are a central element in the ocean ecosystem and nutrient cycles.

The main theme to emerge from the special issue is that, when it comes to mating, copepods exercise considerable control over their own lives. For instance, a laser-illuminated, three-dimensional, video-recording system is being used to study the freshwater species *Cyclops scutifer* (J. R. Strickler). Free-swimming *Cyclops* in an observation chamber were induced to aggregate around a vertical blue laser beam, where they encountered potential mates.

These encounters follow much the same course, as exemplified by observations taking events back to 30 seconds before

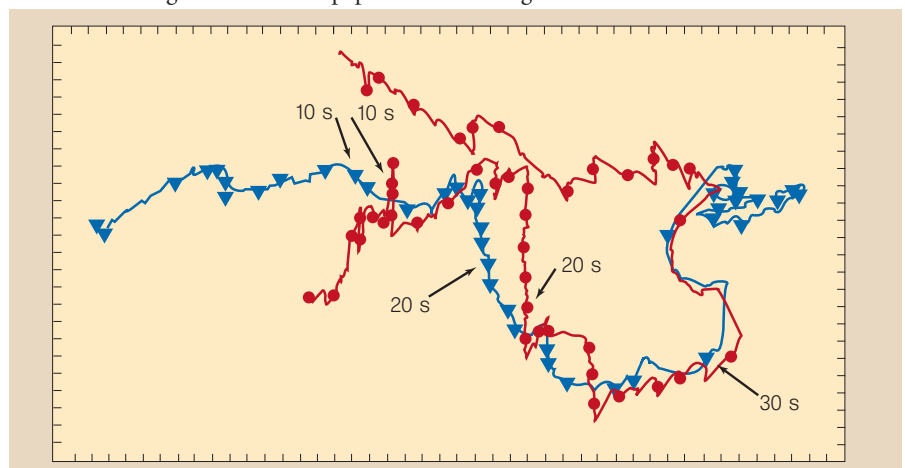


Figure 1 *Cyclops* encounter. This depiction of copepod mating behaviour is J. R. Strickler's work and shows the movement and relative positions of male (blue) and female (red) *Cyclops scutifer*. The male encounters the female at 10 seconds; he follows her for the next 20 seconds, then at 30 seconds they mate. Interval between tick marks on the axes is 1 mm; interval between markers is 1 second. (Revised from Fig. 8 of ref. 4.)

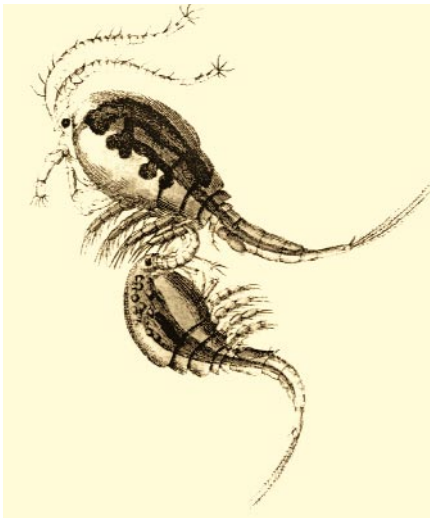


Figure 2 Mating *Cyclops*, as depicted by L. Jurine in an 1820 publication. (Reproduced from ref. 4.)

mating (Fig. 1). No contacts are made in the first ten seconds or so. Once the male perceives a female, he changes his swimming pattern, following the female jump for jump in a synchronized swimming display lasting up to 20 seconds. Once within striking distance, the male lunges towards the female, grasps her tightly, and transfers spermatophores to her reproductive tract. Males, then, actively detect and pursue would-be mates.

Other copepods such as the marine species *Temora longicornis* use different strategies to boost the male–female encounter rate (M. H. Doall *et al.*). *Temora* males swim faster than females, and travel more sinuous routes, often spinning and changing direction abruptly — they seem to be intercepting and tracking hydromechanical ‘footprints’ left in the female wake. This pattern increases the males’ chance of detecting females (although it may also increase their vulnerability to predators).

Females also often ‘hover’ when filter feeding, and males can detect their feeding currents. Whether females actively take part in this ritual by ‘loitering with intent’ is not clear. But tracking moving females is not always successful, and males often backtrack, and cast around before picking up the trail again.

An open question is whether females actively attract males, and whether they reject undesirable partners by trying to escape from them as they would from a predator. A hint that females are indeed active in courtship comes from the work with tethered *Temora* females, which react to chemical exudates of males with little hops distinct from uniform swimming (L. A. van Duren *et al.*). The hydromechanical signals produced by such hops may act as ‘personal ads’, attracting males. Because of the viscous regime, these hydromechanical signals are short-lived, but they greatly increase the ‘encounter volume’ of the female, maximiz-

ing her chance of attracting a male.

From studies of the fine-scale kinematics of the movements of male and female *Temora* comes circumstantial evidence that males may also follow chemical trails left by females (M. J. Weissburg *et al.*). Males zero in on females, and aspects of their behaviour are reminiscent of other animals, such as crabs, lobsters and even ants, that track chemical signals. When searching for a female, the male *Temora* zigzag as if moving in and out of a three-dimensional chemical ‘corridor’; this behaviour is similar to the cross-stream tracking used by male moths when tracing female pheromonal plumes³. Weissburg *et al.* point out that the paired chemosensory organs of the male’s antennae would be ideal for sensing a chemical gradient, and that such pheromonal trails might persist for seconds before the gradient disperses.

Chemical cues may be quite widespread in copepod mating behaviour. Another possible example is that of newly moulted females of *Calanus marshallae*, which sink and apparently leave vertical pheromone trails up to several centimetres long (A. Tsuda and C. B. Miller); males casting out horizontally encounter these trails and ‘waggle’ their way down to the female below. But in both cases, those of *Calanus* and *Temora*, the identity of the presumed

chemical cue is unknown.

A further angle emerges from work with *Tigriopus japonicus* (L. S. Kelly *et al.*). Males grab potential partners, and reject them if, for example, they turn out to be of the wrong species. Mate discrimination seems to depend on contact chemoreception, which is probably mediated by lectin-like glycoproteins on the body surface. Similar signalling may also help prevent interbreeding between the estuarine *Coullana canadensis* and a sibling species where the two overlap off the Florida coast (M. A. Frey *et al.*).

The last words should go to J. Yen and colleagues: “Zooplankton are not aimless wanderers in a featureless environment, an image suggested by the Greek word ‘planktos’. Their ambit is replete with clues that guide them in their search for food or mates and in their other efforts for survival in the ocean.” For all the impressive documentation in the special issue, however, our knowledge of those clues remains preliminary. □

Rory Howlett is deputy biological sciences editor of Nature.

e-mail: nature@nature.com

1. Boxshall, G. A. (ed.) *Mating Biology of Copepod Crustaceans*. Phil. Trans. R. Soc. Lond. B 353, 669–815 (1998).
2. Hardy, A. *The Open Sea: Its Natural History*. I, *The World of Plankton* (Collins, London, 1956).
3. Mafra-Neto, A. & Carde, R. T. *Nature* 369, 142–144 (1994).
4. Strickler, J. R. *Phil. Trans. R. Soc. Lond. B* 353, 671–680 (1998).

Ecology

Polarized flight

Why do mayflies lay their eggs *en masse* on dry asphalt roads? The answer can be found in the latest *Journal of Experimental Biology* (201, 2273–2286; 1998) — but be prepared for a tale of death and deception.

György Kriska and his colleagues observed that every year, in May and June, swarms of mayflies (Ephemeroptera; pictured) mate, not above lakes and rivers, but above dry asphalt roads. The females then land and each lays up to 9,000 eggs, all of which perish.

To investigate this curious behaviour, the authors studied mayfly nuptials on a stretch of asphalt road near Budapest, Hungary. Flies were given a choice of surface on which to lay their eggs — shiny black, white or clear plastic sheets, shiny aluminium foil, matt black or white cloths, and plain asphalt.

The flies unanimously preferred the shiny black surface, to the extent that they even broke off mating if moved to one of the other surfaces. When Kriska *et al.* studied the reflection–polarization properties of these surfaces, they discovered why.

Although reflected light was polarized parallel to the surface of both the black and white plastic sheets, the degree of

polarization was much greater from the black ones. The authors believe that such horizontally polarized light mimics a highly polarized water surface — in other words, the mayflies are deceived into laying their eggs on the road. This ties in with observations that the darker and smoother a region of asphalt (and, hence, the higher the degree of polarization), the more likely the flies are to lay their eggs on it.

This is a real problem. Mayflies could be in danger of extinction as their aquatic habitats are increasingly polluted with herbicides, pesticides and industrial waste. And it’s not just asphalt that fools them. These flies have been seen to lay their eggs on the shiny bodywork and windscreens of cars, in tar pits and even in crude-oil lakes. This trick of the light is not only cruel, it’s deadly.

Alison Mitchell





100 YEARS AGO

The *Lancet* prints the following note on Egyptian native remedies for hydrophobia:— "Though there are no medical accounts of rabies in times past, there are plenty of supposed cures which make it appear as if the disease were well known. ... The modern treatment for a person bitten by a presumably mad dog in Upper Egypt is to kill the dog, extract the spinal cord, bruise the cord with pestle and mortar until a paste is made, and then rub the patient's body all over with paste. Sometimes, too, they burn the dog's hair, and apply the ashes to the bite. The Bedouin make the patient eat the raw liver of the dog, and this is done, too, in the Hausa State of the Western Soudan."

From *Nature* 28 July 1898.

50 YEARS AGO

"Olympic Torches" – The Organising Committee of the XIVth Olympiad approached the Department of Scientific and Industrial Research in 1946 for assistance and advice on the design of a torch to be carried by relays of runners across Europe from the plain of Olympia to Wembley Stadium in London, and on the most suitable fuel. ... After many trials, it was decided that the most suitable fuel was hexamine in the form of tablets. Hexamine gives a non-luminous flame, and to make the flame visible in all weathers 6 per cent of naphthalene was added. ... Prototype torches made at the Fuel Research Station were tried out by runners, first from the South London Harriers and, after certain modifications in design, by officers from the Royal Naval College, Greenwich. In the final design of the torch, seven tablets are enclosed in a perforated metal cylinder with an inner sleeve concealing the lower three tablets. As the upper tablets burn away, the lower reserves are forced up into the burning zone by a spring. In order to facilitate lighting, the Wessex Aircraft Engineering Co., Ltd., Salisbury, provided a tablet of nitrate composition which was introduced on top of the fuel pack.

From *Nature* 31 July 1948.

Many more abstracts like these can be found in *A Bedside Nature: Genius and Eccentricity in Science, 1869–1953*, a 266-page book edited by Walter Gratzer. Contact David Plant, e-mail: subscriptions@nature.com

Vesicular transport

Sticky fingers grab a lipid

Claudia Wiedemann and Shamshad Cockcroft

Inositol lipids first came into the limelight in the 1980s when they were identified as a source of second messengers after cleavage by phospholipase C. In the 1990s they emerged under a different guise — intact inositol lipids were found to interact directly with proteins and to participate in exocytosis (release of messenger molecules from cells), cytoskeletal reorganization and vesicle transport. A report by Simonsen *et al.*¹ on page 494 of this issue now sheds light on how intracellular vesicle transport is connected to inositol lipids. And studies by Gaullier *et al.*² and Patki *et al.*³ (pages 432 and 433), and Bird and Emr⁴ in *Molecular Cell*, feature the protein domain through which this interaction with inositol lipids occurs.

There are several different classes of inositol lipid. The 3-phosphorylated varieties, phosphatidylinositol-3-phosphate (PtdIns(3)P) and the bis- and trisphosphate forms PtdIns(3,5)P₂ and PtdIns(3,4,5)P₃, are formed by phosphoinositol-3-OH kinases (PI3Ks). They cannot be cleaved by phospholipase C, and function as intact lipids. One of them, PtdIns(3,4,5)P₃, takes part in signalling cascades initiated by growth-factor receptors, guanine-nucleotide-binding (G) protein-coupled receptors and cytokine receptors. Another, PtdIns(3)P, has been implicated in intracellular vesicle transport,

and the new reports^{1–4} present a molecular explanation for this interaction.

It was first intimated that PtdIns(3)P might be involved in vesicle transport when the yeast Vps34p protein was identified as a PtdIns3K. Vps34p is the only PtdIns3K to have been found in yeast cells, and it is essential for protein sorting and transport of vesicles from the late Golgi to the vacuole (the equivalent of the mammalian lysosome; Fig. 1). In mammalian cells, PtdIns(3)P is generated by different PtdIns3K isoforms and, like PtdIns(3)P in yeast, it is required for transport from the *trans*-Golgi network to the lysosomes. It is also involved in the endocytic pathway in mammals, whereby plasma-membrane receptors are internalized and extracellular material is taken up by the cell.

Endocytosis requires a small GTP-binding protein called Rab5 (Fig. 1). Early studies established a connection between Rab5, PtdIns3K and the early endosomal autoantigen EEA1, which is associated with early endosomes in mammalian cells. Now we have evidence of how this connection occurs (Fig. 2). Gaullier *et al.*² and Patki *et al.*³ have identified a protein domain in EEA1, the FYVE finger module, which binds to PtdIns(3)P. This domain was christened the FYVE finger based on the first letters of four proteins that contain it. The motif has been

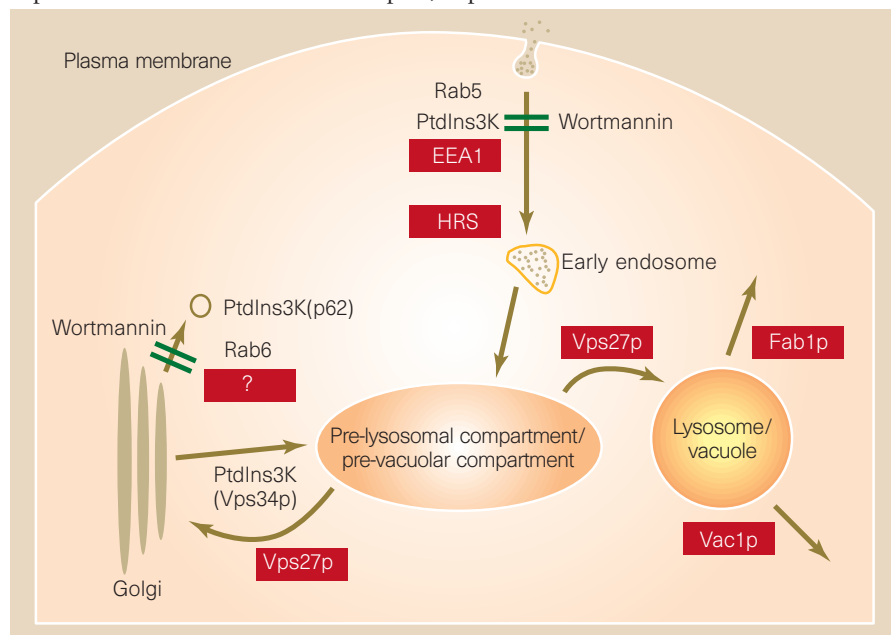


Figure 1 Membrane-transport events that rely on proteins containing the FYVE finger domain. (Red boxes indicate proteins with FYVE domains.) Early endosomal autoantigen (EEA1) and Hrs are associated with early endosomes in mammalian cells¹³; Fab1p, Vac1p and Vps27p are implicated in vacuolar transport in yeast. Fab1p is a putative phosphoinositide-5-OH kinase that is probably responsible for converting¹⁶ PtdIns(3)P to PtdIns(3,5)P₂ and for traffic out of the vacuole. Vps27p and, possibly, Vac1p are components of the prevacuolar endosome, which is required for protein sorting to the vacuole⁷. Vac1p is required for vacuole inheritance by the daughter cell during budding^{9,10}.

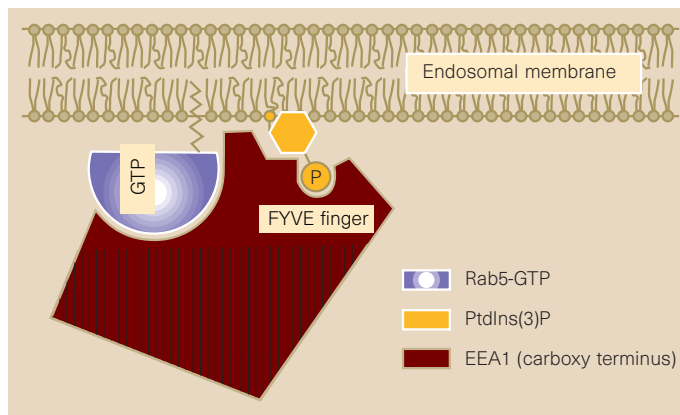


Figure 2 Model for the association of EEA1 with the endosomal membrane, based on the results of four new studies¹⁻⁴. The carboxy terminus of EEA1 contains adjacent binding sites for Rab5-GTP (amino-acid residues 1,277–1,348) and PtdIns(3)P (residues 1,349–1,411).

found in at least 29 proteins from yeast, worms, plants, insects and mammals. It is the first protein domain known to bind this inositol lipid.

The FYVE domain is about 70 amino acids long, has eight potential Zn²⁺-coordinating cysteine residues, and binds two molecules of zinc. It was originally identified after observations that wortmannin — an inhibitor of many PI3Ks — caused the redistribution of EEA1 from the endosome to the cytosol^{5,6}. The FYVE motif is located at the carboxy terminus of EEA1, and it is homologous to domains found in Vps27p, Fab1p, Vac1p (yeast proteins implicated in vesicular transport)⁷⁻¹⁰, Hrs (a mammalian endosomal protein)¹¹ and Hrs-2 (another mammalian protein involved in calcium-regulated exocytosis)¹².

Gaullier *et al.*² and Patki *et al.*³ tie these findings together and show that EEA1 binds to PtdIns(3)P through its FYVE domain. They find that both wortmannin and mutations in the FYVE domain that disrupt binding to PtdIns(3)P cause dissociation and redistribution of EEA1 to the cytosol. Burd and Emr⁴ further show that Vps27p, Fab1p and Vac1p bind PtdIns(3)P *in vitro*. And here comes the surprise — Simonsen *et al.*¹ show that the requirement for PtdIns(3)P can be bypassed if there is an excess of membrane-bound Rab5-GTP. It turns out that EEA1 not only binds PtdIns(3)P through the FYVE finger domain, but that it also interacts with Rab5-GTP through an adjacent domain in its carboxy terminus (Fig. 2). Thus, under physiological conditions, EEA1 is maintained at the membranes by Rab5-GTP and PtdIns(3)P.

The requirement for EEA1 in endocytic fusion has been shown by Mills *et al.*¹³ in another recent paper. If they depleted EEA1 from membranes and cytosol, they found that fusion was blocked. Moreover, when they added the carboxy-terminal domain of EEA1 only (residues 1,098–1,411), they observed the same effect. These residues encompass the FYVE domain, as well as the site for Rab5 binding (Fig. 2).

The observation that a single protein (EEA1) can bind both lipid (PtdIns(3)P) and protein (Rab5) provides a unique

method of targeting to endosomes only. Other organelles, such as the *trans*-Golgi network, contain PtdIns(3)P. But, because Rab5 is found only in the early endosome, EEA1 cannot bind these organelles. So, EEA1 provides directionality to Rab5-dependent endocytic fusion. Rab proteins are involved in many other fusion reactions; for example, Rab3 functions in exocytosis, where an Hrs isoform (Hrs-2) is also involved¹². Moreover, Rab6 and p62, a PtdIns3K^{11,14}, are required for budding of exocytic vesicles from the *trans*-Golgi network in mammalian cells. This concept of using a protein with dual binding specificity is an elegant device for introducing specificity and directionality to

Drug delivery

A lesson from secretory granules

Ronald A. Siegel

A primary task in formulating drug-delivery systems is to package the drug molecules at high concentration, and release them later according to a predetermined temporal and spatial programme. In the body, this task is regularly carried out by secretory granules, and on page 459 of this issue¹ Needham and colleagues describe how they have learned from that example. By a clever combination of polymer-gel science and lipid chemistry, they have constructed a granule mimic that can store a drug for a desired period and rapidly release it when properly stimulated.

Hydrogels (crosslinked polymer networks which, in the right conditions, swell in water) have long been studied and employed as drug-delivery vehicles². By physically trapping a drug in a hydrogel and using the sieving properties of the matrix to retard molecular movement, or simply by maintaining the gel in a dry state, the drug can be retained inside the device for long periods. Covalent or noncovalent bonding of drugs to side-chains of the polymer gel is another way of storing a drug and releasing it under controlled conditions. Researchers have sought ways to effect release and modulate release rates by using changes in pH or tem-

perature to alter the degree and rate of gel swelling; similarly, external stimuli can break bonds between drug and polymer, allowing the drug to be released. Conventional hydrogel-based delivery systems release their contents over periods of hours to days, but this timescale can be reduced by making the gels smaller. In most cases the time required for complete swelling and release of the gel contents varies linearly to quadratically with gel diameter, depending on the rate-controlling process^{2,3}. So a tenfold reduction in diameter can lead to up to a hundredfold increase in speed.

Although most hydrogel drug-delivery constructs are made with artificial polymers or reconstituted polysaccharides or proteins, it has not escaped notice that nature itself has evolved a very efficient molecular storage and delivery device — the secretory granule⁴.

For example, mast cells can store histamine at high concentrations for indefinite periods in granules, and release it within milliseconds after the appropriate electrochemical stimulation. The structure of the mast-cell secretory granule — a dense network of sulphated proteoglycans, around a micrometre in size — allows it to condense in the presence of positively charged counterions such as

1. Simonsen, A. *et al.* *Nature* **394**, 494–498 (1998).
2. Gaullier, J.-M. *et al.* *Nature* **394**, 432–433 (1998).
3. Patki, V., Lawe, D. C., Corvera, S., Virbasius, J. V. & Chawla, A. *Nature* **394**, 433–434 (1998).
4. Burd, C. G. & Emr, S. D. *Mol. Cell* **2**, 157–162 (1998).
5. Stenmark, H., Aasland, R., Toh, B. & D'Arrigo, A. *J. Biol. Chem.* **271**, 24048–24054 (1996).
6. Patki, V. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 7326–7330 (1997).
7. Piper, R. C., Cooper, A. A., Yang, H. & Stevens, T. H. *J. Cell Biol.* **131**, 603–617 (1995).
8. Yamamoto, A. *et al.* *Mol. Biol. Cell* **6**, 525–539 (1995).
9. Weisman, L. S. & Wickner, W. *J. Biol. Chem.* **267**, 618–623 (1992).
10. Conradt, B., Shaw, J., Vida, T., Emr, S. & Wickner, W. *J. Cell Biol.* **119**, 1469–1479 (1992).
11. Bean, A. J., Seifert, R., Chen, Y. A., Sacks, R. & Scheller, R. H. *Nature* **385**, 826–829 (1997).
12. Jones, S. M., Crosby, J. R., Salamero, J. & Howell, K. E. *J. Cell Biol.* **122**, 775–788 (1993).
13. Mills, I. G., Jones, A. T. & Clague, M. J. *Curr. Biol.* **8**, 881–884 (1998).
14. Jones, S. M. & Howell, K. E. *J. Cell Biol.* **139**, 339–349 (1997).
15. Komada, M., Masaki, R., Yamamoto, A. & Kitamura, N. *J. Biol. Chem.* **272**, 20538–20544 (1997).
16. Dove, S. K. *et al.* *Nature* **390**, 187–196 (1997).

histamine, serotonin and calcium. These ions bind tightly to and neutralize the fixed charge groups, causing the granule network to shrink. Shrinking is favoured when the counterion is polyvalent or partially hydrophobic.

If a secretory granule consisted only of a charged network, it would not hold its payload very long in the intracellular or the extracellular environment. Sodium and potassium ions would invade the granules and rapidly 'exchange out' the incorporated species. Because these metal ions do not bind to the fixed negative-charge groups, they would establish an osmotic pressure inside the granule, causing it to swell and release its contents. To prevent this, the granule is coated with a lipid membrane that blocks ion transport. Secretion of the stored contents requires electrochemically stimulated fusion of the granule's membrane with the cell membrane, which exposes the matrix to the extracellular medium, triggering ion exchange and swelling.

If nature can do this, why can't we? That was the question asked by Needham and colleagues, and their response is outlined in Fig. 2 of the paper on page 459. Using technology that had already been developed⁵, they synthesized crosslinked polymethacrylic-acid microgels that had a swollen diameter of 6.5 mm at pH 7.4. They then incorporated the hydrophobic, cationic, anticancer drug doxorubicin hydrochloride at pH 5.0, which neutralized the acid groups, causing the gel to shrink. As a final step, they used a newly devised process to coat the collapsed gel with a lipid bilayer, and showed that the lipid was present exclusively at the surface. The lipid coating all but prevents dissipation of pH gradients, and the acidified construct remains stable in pH 7.4 phosphate-buffered saline, a proxy for body fluids, for at least 48 hours. By this means they mimic the storage aspect of a granule.

To demonstrate 'quick release', the authors exposed microspheres in phosphate-buffered saline at pH 7.4 to electroporation fields that were strong enough to breach the lipid coatings. A brief electroporation pulse led to rapid swelling of the gels which was complete within seconds, and all the incorporated doxorubicin was released within minutes. Although these processes take somewhat longer than they do in nature, the times concerned are more than acceptable for many drug-release purposes.

How might such a system be used in practice? The particles are of such size that they will be rapidly cleared by the reticuloendothelial system, which consists of scavenger cells that continually patrol the body. So, without modification, the system is probably best suited for local administration, be it subcutaneous, intramuscular or intraperitoneal. The authors suggest, however, that the synthesis can be altered to make much smaller particles that will avoid the reticuloendothelial system

and can also leak through the porous capillary walls of tumours, providing targeted delivery of anticancer agents. The practicality of the system might be further improved by tethering certain molecules, ones which bind specifically or nonspecifically to target cells or extracellular matrix, to the lipid coating, thus localizing the microspheres at a particular site and perhaps preventing side-effects. Collagen, peptides containing arginine-glycine-aspartate sequences and Fab fragments of antibodies are logical candidates. This strategy has been investigated with liposomes, another type of drug carrier⁶.

Needham and colleagues' prototype system triggers drug release by electroporation, which may be difficult to effect *in situ*, and the authors point out that modifications to the lipid coating may make it sensitive to stimuli such as temperature or ultrasound. Work with liposomes also points to a second, potentially powerful, refinement of the system^{7,8}. By anchoring suitable stimulus-sensitive polymers in the membrane, one can destabilize the lipid coating, and therefore trigger drug release, by changes in local temperature, pH and glucose concentration, or by illumination at a particular wavelength. Even more sophisticated systems can be envisaged in which lipid-coated, bioadhesive microgels that respond to different stimuli, each type of microgel containing a different agent, are mixed together and delivered locally or regionally. Such a system could permit localized combination therapies, in which the delivery of the different agents occurs according to a predetermined sequence.

These kinds of elaborations would add further complexity to Needham and colleagues' system — the combinatorial possibilities are numerous, and would require a great deal of further development and testing. Moreover, the effectiveness of the original concept needs to be tried out in an animal model. At the least, however, the authors have succeeded in showing how some cunning chemistry can be used to emulate a physiological process for the purposes of improved drug delivery. □

Ronald A. Siegel is in the Departments of Biopharmaceutical Sciences and of Pharmaceutical Chemistry, School of Pharmacy S-926, University of California at San Francisco, San Francisco, California 94143-0446, USA.
e-mail: ras@rebus.ucsf.edu

1. Kiser, P. F., Wilson, G. & Needham, D. *Nature* **394**, 459–462 (1998).
2. Langer, R. S. & Peppas, N. A. *Biomaterials* **20**, 201–214 (1981).
3. Tanaka, T. & Fillmore, D. J. *Chem. Phys.* **70**, 1214–1218 (1979).
4. Fernandez, J. M., Villalon, M. & Verdugo, P. *Biophys. J.* **59**, 1022–1027 (1991).
5. Kashiwabara, M., Fujimoto, K. & Kawaguchi, H. *Coll. Polym. Sci.* **273**, 339–345 (1995).
6. Yerushalmi, N. & Margalit, R. *Archiv. Biochem. Biophys.* **349**, 21–26 (1998).
7. You, H. & Tirrell, D. J. *Am. Chem. Soc.* **113**, 4022–4023 (1991).
8. Meyer, O., Papahadjopoulos, D. & Leroux, J. *FEBS Lett.* **421**, 61–64 (1998).

Daedalus

A chance of justice

A legal verdict, says Daedalus, pretends to be a certainty; but in truth it is merely a probability. Consider, for example, those plaintiffs who claim to have been made ill by silicone breast-implants, radioactivity or tobacco. Clearly it is impossible to be sure that they would not have got just as ill if they had never encountered these perils.

So Daedalus is musing on the character and process of a truly scientific court. If, for example, epidemiological evidence shows that passive smoking in the workplace increases your chance of getting lung cancer by 1%, then the judge could decide the case won if the number 100 turns up on his random number generator, but lost if 1–99 turn up. Chancy or vexatious litigants would be strongly discouraged.

It might be fairer, however, to allow an openly statistical verdict. A defendant found 'probably guilty' (perhaps set at 91.7% if 11 jurors out of 12 reckon he did it) might incur a fine or punishment reduced in proportion. A defendant found 'probably innocent' (say, 8.3% or 16.7%) might leave the court unpunished, but with a definite 'stain' on his character. Thereafter, until the stain was declared spent, or expunged by good behaviour, he would be literally 'a suspicious character' — which would tell against him if he came up again on a similar charge.

This approach would fit well into British society. British motorists already accumulate 'stains' on their driving licences for each small offence. Enough staining cancels the licence. The British honours system awards a 'shine' on the character of good eggs and praiseworthy types; if they later go to the bad, the shine can be withdrawn again. Affirmative action gives whole groups a collective shine, entitling them to jobs, presumptions of virtue or innocence, and so on.

Such an open system of honours and dishonours is much fairer than the 'dossier societies' run by dictators. Secret dossiers, for some reason possibly connected with the second law of thermodynamics, only accumulate evidence against their subjects. Only those faceless apparatchiks who never put a foot wrong or do anything original, flourish under them. Ominously, dossiers are now growing fast even in the democracies, in the form of credit ratings, referees' reports, compulsory secret reporting of 'suspicious' bank deposits, and so on. An open, numerical 'stains and shines' system might just stop the rot.

David Jones

Kroto and charisma

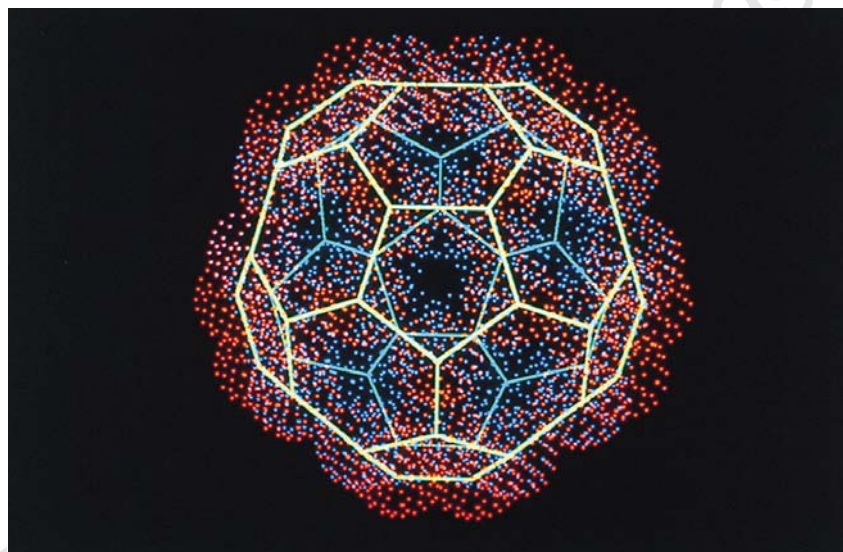
A vision of a dome led to the naming of buckminsterfullerene. This carbon cluster has become an icon for chemistry, thanks to the media-friendly appeal it shares with co-discoverer Sir Harry Kroto.

Martin Kemp

Sometimes it is difficult to know why particular scientific discoveries hit the public headlines. The hugely technical solution to Fermat's last theorem hardly seemed to possess the ingredients for a hit. It could not even resort to the kind of visual appeal that normally helps to bring a discovery before the public eye. By contrast, the popular impact of the 'buckyball', C_{60} , unabashedly relied upon visual charisma.

The story of the identification, modelling and naming of buckminsterfullerene in 1985 already has the quality of legend — like August Kekulé's vision of the benzene ring in a dream about a snake biting its tail. Sir Harry Kroto tells how his work at Rice University in Houston, Texas (with a team including Robert Curl and Richard Smalley), led to the identification of the 60-atom cluster of carbon which exhibited a stability at odds with any graphite-like or diamond-like configuration. An architecture was required that closed off the apparently dangling valencies in any of the immediately plausible arrangements.

It was difficult to see how a closed polygon of hexagons, like a ball of graphite, could work. The key moment came with the introduction of pentagonal faces (prefigured by Eiji Osawa in 1970), which can effect closure on extended orbs of hexagons of varied dimensions.



Molecular model of buckminsterfullerene in galactic guise.

The suggestion welled up from Kroto's memory of the geodesic dome designed by the architect and visionary inventor, Buckminster Fuller, for the American pavilion at Expo '67 in Montreal. Kroto also recalled making a Fuller-style cardboard skymap for his children in the form of a "stardome". The stardome comprised a truncated icosahedron of 20 hexagonal and 12 pentagonal faces, with 60 vertices.

On this hunch that pentagons were involved, Smalley worked into the night to

construct the first paper model of the hypothetical structure.

A crucial component that Kroto brought to the buckyball team was his natural instinct as a designer. Indeed, he had long hankered after a career in graphic design, and, before the consuming success of C_{60} , planned to found a studio for scientific graphics. He is one of those scientists, like Leonardo and Kepler, naturally drawn to the tangible beauty of complex symmetries in works of nature and art.

Compared to the initial announcement of C_{60} and the technical outline of "our somewhat speculative structure", which occupied less than 1.5 columns of dryly laconic prose in *Nature*, the public broadcasting of the victory of the buckyball team was marked by a hatful of metaphorical goals. Extolling the "cosmic and microcosmic charisma of the soccerball", Kroto's 1996 Nobel prize address expressed delight with its ascent to stardom: "This elegant molecule ... has fascinated scientists, delighted lay people, and has infected children with a new enthusiasm for science. And in particular it has given chemistry a new lease of life."

At present it looks as if the buckyball is to become the kind of modern icon for chemistry that DNA has become for molecular biology. It also helps, of course, if one or more of the discoverers exhibit a public charisma to match that of the molecule. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk



The five-a-side team of buckyball discoverers: (from left) Sean O'Brien, Richard Smalley, Robert Curl, Harry Kroto and James Heath.

Ploughing up the wood-wide web?

Key species groups that affect major ecological processes are vital components of community diversity. Many such key groups are found in the soil, including the mycorrhizal fungi that may connect plants into a functional "wood-wide web"¹. Arbuscular mycorrhizal associations are formed by fungi of the order Glomales with 90% of land plant families, and many arbuscular mycorrhizal fungi are thought to have a broad host range². Here we show that, despite this broad host range, the diversity of arbuscular mycorrhizal fungi is strikingly low in arable sites compared with a woodland.

The arbuscular mycorrhizal fungi that colonize roots cannot be reliably identified below the genus level except by molecular methods. We examined roots from five abundant woodland plant species at four sites within a broadleaved wood dominated by oak (*Quercus petraea*, colonized by ecto-

mycorrhizal fungi), and sycamore (*Acer pseudoplatanus*, colonized by arbuscular mycorrhizal fungi), at Castle Howard, North Yorkshire, UK. Partial fungal small-subunit ribosomal RNA sequences were amplified, cloned and screened for differences in restriction pattern by restriction-fragment length polymorphism (RFLP). We sequenced selected clones to determine their phylogenetic position (Fig. 1). For comparison, we sampled pea, maize and wheat crops on three farms within a 55-km radius of the woodland site.

There are three families in the Glomales, represented in our samples by the genera *Glomus*, *Acaulospora* and *Scutellospora*, and these are readily distinguished as distinct sequence clusters (Fig. 1). Morphological studies confirmed that these three genera were present in these roots³. In arable sites, 92% of sequences represented *Glomus mosseae* or closely related taxa, whereas

those from woodland were much more diverse (Fig. 1). This was true even though the arable sites were separated by up to 66 km and three host species were sampled. The combined woodland samples had a much higher diversity of RFLP types (Shannon–Weiner $H=0.144$) than the combined arable samples ($H=0.398$). In both wood and field, we often obtained identical sequences from different plant species, suggesting that the broad host range exhibited by many cultured arbuscular mycorrhizal fungi² may also be realized in nature.

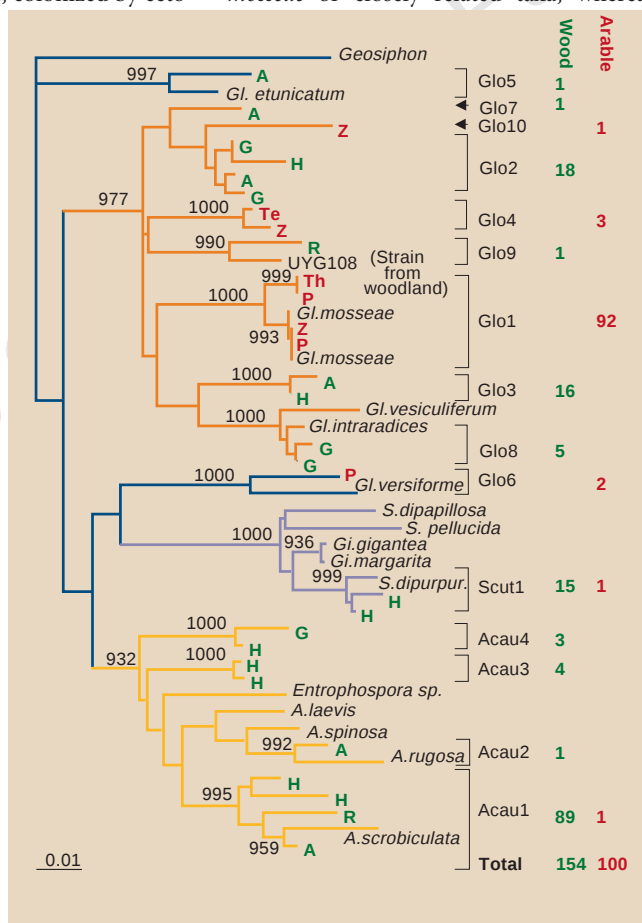
Given the broad host range of some arbuscular mycorrhizal fungal taxa, the change in sequence composition and low diversity of the fungi in arable fields is probably not a result of plant monoculture *per se*, but reflects other aspects of the agronomic regime such as ploughing, fertilization or fungicide application. In all the arable fields, regardless of host plant or location, the dominant arbuscular mycorrhizal fungal type was a putative *G. mosseae* not found in the woodland. This species sporulates abundantly and colonizes readily from spores, which may be more important in a field that is ploughed annually than in woodland⁴.

Arbuscular mycorrhizal fungi differ widely in their biological properties, and presumably have several different roles in ecosystems⁵. The low taxonomic diversity of arbuscular mycorrhizal fungi in arable fields indicates that their functional contribution may be less there than in woodland. It has been suggested that low ecosystem diversity may be associated with impaired function^{6,7} and reliability^{8,9}. Our results show that microbes need to be considered in any assessment of the effects of agriculture on biological diversity and that intensive arable agriculture may be operating at minimum levels of diversity for at least one key functional group.

T. Helgason, T. J. Daniell, R. Husband, A. H. Fitter, J. P. W. Young

Biology Department, University of York,
PO Box 373, York YO10 5YW, UK
e-mail: th7@york.ac.uk

Figure 1 Neighbour-joining phylogeny of arbuscular mycorrhizal fungal DNA sequences amplified from plant roots sampled at woodland (green letters) and arable (red letters) sites. Fungal sequences are identified by the host plant from which they were isolated. A, *Ajuga*; H, *Hyacinthoides*; E, *Epilobium*; G, *Glechoma*; R, *Rubus* (from the woodland site, sampled in July 1996). P, *Pisum* and Te, *Triticum* from Escrick; Th, *Triticum* from High Mowthorpe; Z, *Zea* from Bedale (arable sites, sampled at three time points during 1997). Orange branches correspond to the arbuscular mycorrhizal fungus family Glomaceae, purple branches to Gigasporaceae, yellow branches to Acaulosporaceae, and dark blue branches to the two taxa that do not cluster with any of these families. Partial small-subunit



ribosomal DNA fragments of about 550 base pairs were amplified using *Pfu* DNA polymerase and primers NS31 (ref. 10) and AM1 (5'-GTT TOC CGT AAG GCG CCG AA-3', designed to amplify fungal and exclude plant DNA sequences). Cloned products were digested with *HinfI* and *AluI* and selected samples sequenced. Named sequences are from GenBank or from library cultures sequenced in our laboratory. Bootstrap values of >90% are shown at the nodes. RFLP types defined for analysis are shown on the right (Glo, *Glomus*; Acau, *Acaulospora*; Scut, *Scutellospora*), along with the number of clones of each type found in woodland (green numbers) and arable land (red numbers). Previously unpublished sequences have been deposited in GenBank under accession numbers AF074340–AF074373.

- Simard, S. W. *et al.* *Nature* **388**, 579–582 (1997).
- Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* (Academic, San Diego, 1996).
- Merryweather, J. & Fitter, A. H. *New Phytol.* **138**, 117–129 (1998).
- Clapp, J. P., Young, J. P. W., Merryweather, J. W. & Fitter, A. H. *New Phytol.* **130**, 259–265 (1995).
- Newsham, K. K., Fitter, A. H. & Watkinson, A. R. *J. Ecol.* **83**, 991–1000 (1995).
- Chapin, F. S. *et al.* *Science* **277**, 500–504 (1997).
- Tilman, D. *et al.* *Science* **277**, 1300–1302 (1997).
- McGrady-Steed, J., Harris, P. M. & Morin, P. J. *Nature* **390**, 162–165 (1997).
- Naeem, S. & Li, S. B. *Nature* **390**, 507–509 (1997).
- Simon, L., Lalonde, M. & Bruns, T. D. *Appl. Environ. Biol.* **58**, 291–295 (1992).

FYVE fingers bind PtdIns(3)P

The membrane lipid phosphatidylinositol-3-phosphate (PtdIns(3)P) is constitutively produced by yeast and higher eukaryotes through the phosphorylation of phosphatidylinositol by phosphatidylinositol-3-OH kinase¹ (PI(3)K). PtdIns(3)P is important for vesicular transport¹, but little is known about how it acts, and proteins that specifically recognize it have not yet been identified. Here we identify the FYVE finger², an evolutionarily conserved double-zinc-binding domain (see Supplementary information), as a protein structure that binds to PtdIns(3)P with high specificity.

The human early-endosomal autoantigen EEA1 (ref. 3) is regulated by PI(3)K activity⁴. Its carboxy terminus (EEA1-CT) contains a FYVE finger and co-localizes with the GTPase Rab5 (ref. 3) on early endosomes, as shown by confocal immunofluorescence microscopy of transfected baby hamster kidney (BHK) cells (Fig. 1a, b). In contrast, we detected no membrane-associated EEA1-CT in cells treated with low concentrations of the PI(3)K inhibitors wortmannin (Fig. 1c, d) and LY294002 (not shown).

The expression of a zinc-binding-defective mutant EEA1-CT protein (C1405S; cysteine 1,405 is replaced with serine²) resulted in a similar effect to that of wortmannin treatment² (Fig. 1e, f). This indicates that the association of EEA1 with endosomes could require the interaction of the FYVE finger with a 3'-phosphoinositide. However, as the FYVE finger alone does not bind to membranes², another domain of EEA1-CT must also participate in the membrane association⁵.

To determine whether FYVE fingers from structurally unrelated proteins can functionally replace each other, we substituted the FYVE finger of EEA1-CT with that of the tyrosine-phosphorylated endosomal protein Hrs⁶. We expressed the resulting hybrid protein in BHK cells. Confocal immunofluorescence microscopy of these cells showed that the hybrid protein was abundant on Rab5-positive vesicular structures (Fig. 1g, h), indicating that the Hrs FYVE finger could indeed replace the EEA1 FYVE finger with respect to endosomal targeting of EEA1. As with EEA1-CT, the membrane localization of the hybrid construct was also sensitive to wortmannin (Fig. 1i, j), LY294002 (not shown), and to a FYVE-domain mutation affecting zinc binding (C215S; Fig. 1k, l). These data provide *in vivo* evidence that the FYVE finger is a conserved 3'-phosphoinositide-interacting domain.

We next prepared glutathione S-trans-

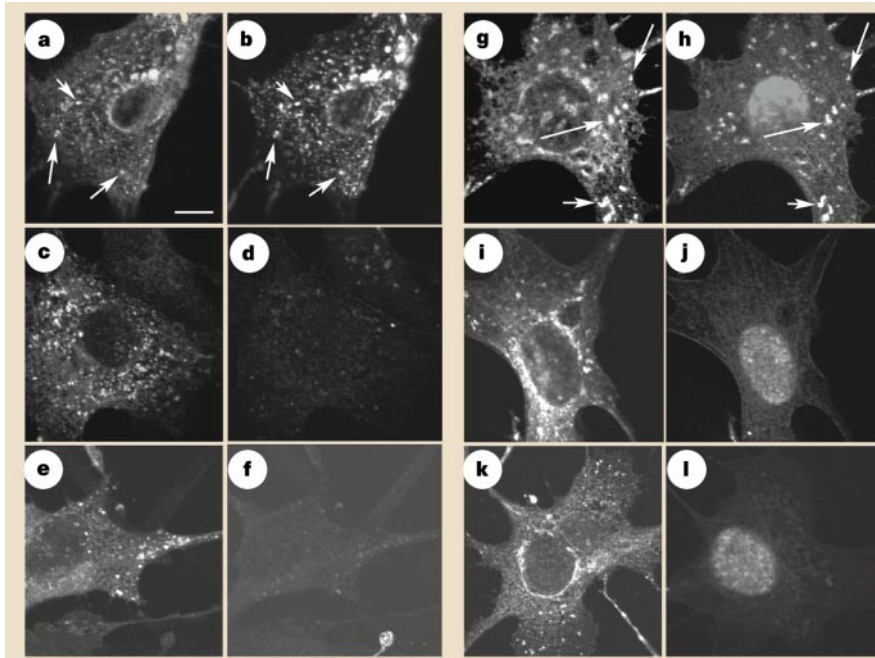


Figure 1 The endosomal localization of EEA1-CT and its Hrs FYVE-finger hybrid depends on PI(3)K activity and an intact FYVE finger. BHK21 cells were co-transfected with Rab5 and Myc-epitope-tagged EEA1₁₂₅₇₋₁₄₁₁ (a-d), Myc-EEA1₁₂₅₇₋₁₄₁₁^{C1405S} (e-f), Myc-EEA1₁₂₅₇₋₁₃₄₉/Hrs₁₆₁₋₂₂₁ (g-j) and Myc-EEA1₁₂₅₇₋₁₂₄₉/Hrs₁₆₁₋₂₂₃^{C215S} (k, l) and then incubated for 20 minutes at 37 °C in the absence (a, b, e-h, k, l) or presence (c, d and i, j) of 20 nM wortmannin. The cells were permeabilized with 0.05% saponin to wash out cytosolic protein, then fixed and finally stained with anti-Rab5 (a, c, e, g, i, k) or anti-Myc-epitope (b, d, f, h, j, l) antibodies and examined by confocal immunofluorescence microscopy. a shows the same cell as b, c shows the same cell as d, and so on. Arrows indicate examples of co-localization of Rab5 with epitope-tagged protein. Scale bar, 10 μm.

ferase (GST) fusion proteins containing the EEA1 or Hrs FYVE fingers and measured the binding of the immobilized fusion proteins to radiolabelled liposomes⁷. Neither protein bound to the liposome matrix, which was composed of 63–65% phosphatidylcholine, 20% phosphatidylserine and 15% phosphatidylethanolamine (Fig. 2a).

In contrast, both FYVE-finger fusion proteins bound strongly to liposomes containing 2% PtdIns(3)P. This binding was highly specific, as the FYVE fusion proteins did not bind to liposomes containing similar amounts of a range of phosphoinositides (Fig. 2a).

Neither GST alone nor GST fused with the zinc finger of Rabphilin-3A (this zinc finger is related to but distinct from the FYVE finger^{2,8}) bound to PtdIns(3)P. This was also true for zinc-binding-defective FYVE-finger mutants of EEA1 and Hrs (Fig. 2b). Pretreatment of the wild-type GST-EEA1-FYVE construct with EDTA to remove Zn²⁺ strongly reduced its ability to bind to PtdIns(3)P. These results show that the EEA1 and Hrs FYVE fingers bind directly and specifically to PtdIns(3)P, and that bound Zn²⁺ is needed for this activity.

Using a different lipid-binding assay, Patki *et al.*⁹ obtained similar results with EEA1. Thus the binding of FYVE fingers to PtdIns(3)P seems to be affected little by the overall lipid composition of the membrane.

The FYVE-finger motif is highly con-

served in more than 30 different proteins of yeast, nematode, plant, insect and mammalian origin (see Supplementary information). In addition to the mammalian proteins EEA1 and Hrs, we have recently found that a *Schizosaccharomyces pombe* FYVE-finger protein, AC19A8.05C, binds PtdIns(3)P (J.-M. G. and H. S., unpublished data). Thus, not only the structure but also the function of FYVE fingers is evolutionarily conserved.

A zinc-binding-deficient EEA1 FYVE-finger mutant lacks the ability of the wild-type protein to regulate endocytic membrane fusion⁵ and similar mutations in the FYVE fingers of the yeast proteins Vac1p and Vps27p strongly impair their ability to sustain membrane traffic to the vacuole^{10,11}. This constitutes compelling functional evidence that FYVE fingers are essential structural elements in a subset of proteins regulating endocytic/vacuolar membrane traffic. Our results offer a plausible explanation for the importance of PtdIns(3)P and FYVE domains in these processes by implicating FYVE-finger proteins as effectors of PtdIns(3)P. The finding that the FYVE fingers tested bind to PtdIns(3)P and not to phosphatidylinositol-3,5-bisphosphate (Fig. 2a) argues against the idea that the only role of PtdIns(3)P in membrane traffic could be as a precursor of phosphatidylinositol-3,5-bisphosphate (ref. 12). It will be interesting to

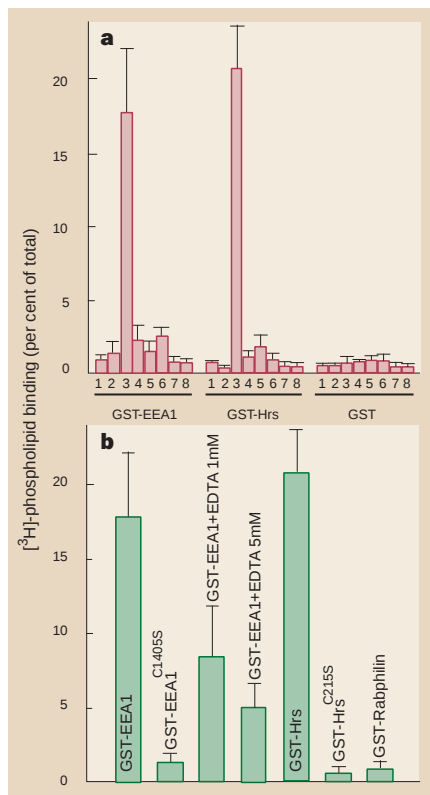


Figure 2 FYVE domains bind specifically to PtdIns(3)P *in vitro*. Portions (0.5 nmol) of GST alone or fused to EEA1¹³⁰⁷⁻¹⁴¹¹, Hrs¹⁻²⁸⁹, EEA1¹³⁰⁷⁻¹⁴¹¹ C^{140SS}, Hrs¹⁻²⁸⁹ C^{215S} or Rabphilin⁴⁵⁻¹⁸² were immobilized on glutathione-Sepharose beads, and then incubated in the presence of [³H]phosphatidylcholine-containing liposomes (0.35 mg ml⁻¹) containing 2% of: **a**, the indicated inositol lipids (see below) or **b**, PtdIns(3)P. Where indicated, the immobilized protein had been preincubated with EDTA. All values represent the mean $\pm$ s.e.m. of three experiments performed in triplicate. 1, phosphatidylcholine; 2, phosphatidylinositol; 3, PtdIns(3)P; 4, phosphatidylinositol-3,4-bisphosphate; 5, phosphatidylinositol-3,5-bisphosphate; 6, phosphatidylinositol-3,4,5-trisphosphate; 7, phosphatidylinositol-4-phosphate; 8, phosphatidylinositol-4,5-bisphosphate.

learn whether PtdIns(3)P and FYVE fingers have other cellular functions as well.

Jean-Michel Gaullier, Anne Simonsen, Antonello D'Arrigo, Bjørn Bremnes, Harald Stenmark

Department of Biochemistry,
The Norwegian Radium Hospital,
N-0310 Oslo, Norway
e-mail: stenmark@ulrik.uio.no

Rein Aasland

Department of Molecular Biology, University of
Bergen, N-5020 Bergen, Norway

1. Toker, A. & Cantley, L. C. *Nature* **387**, 673–676 (1997).
2. Stenmark, H., Aasland, R., Toh, B. H. & D'Arrigo, A. *J. Biol. Chem.* **271**, 24048–24054 (1996).
3. Mu, F. T. et al. *J. Biol. Chem.* **270**, 13503–13511 (1995).
4. Patki, V., Virbasius, J., Lane, W. S., Toh, B. H., Shpetner, H. S. & Corvera, S. *Proc. Natl Acad. Sci. USA* **94**, 7326–7330 (1997).
5. Simonsen, A. et al. *Nature* **394**, 494–498 (1998).
6. Komada, M., Masaki, R., Yamamoto, A. & Kitamura, N. *J. Biol. Chem.* **272**, 20538–20544 (1997).

7. Schiavo, G., Gu, Q.-M., Prestwich, G. D., Söllner, T. H. & Rothman, J. E. *Proc. Natl Acad. Sci. USA* **93**, 13327–13332 (1996).
8. Shirataki, H. et al. *Mol. Cell. Biol.* **13**, 2061–2068 (1993).
9. Patki, V., Virbasius, J. V., Chawla, A., Lawe, D. C. & Corvera, S. *Nature* **394**, 433–434 (1998).
10. Piper, R. C., Cooper, A. A., Yang, H. & Stevens, T. H. *J. Cell Biol.* **131**, 603–617 (1995).
11. Burd, C. G., Peterson, M., Cowles, C. R. & Emr, S. D. *Mol. Biol. Cell* **8**, 1089–1104 (1997).
12. Dove, S. K. et al. *Nature* **390**, 187–192 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

A functional PtdIns(3)P-binding motif

Treating cells with the phosphatidylinositol-3-OH kinase (PI(3)K) inhibitor wortmannin causes the dissociation of the early-endosomal antigen EEA1 from early endosomes¹. EEA1 from cytosolic extracts binds to liposomes containing phosphatidylinositol-3-phosphate (PtdIns(3)P), the major product of PI(3)K in yeast and mammalian cells^{1,2}. Here we show that a RING zinc-finger domain at the carboxy terminus of EEA1, previously identified and named the 'FYVE' domain³, binds directly and specifically to PtdIns(3)P. This indicates that proteins containing this motif

may be downstream effectors of PI(3)K in yeast and mammalian cells.

The RING finger of EEA1 exhibits extensive homology to domains in Vac1p and Vps27p, two proteins that may be involved in the Vps34p-dependent vacuolar-protein-sorting pathway in yeast^{4,5}, and in Hrs-2 and FGD-1, proteins involved in signal transduction and in actin-cytoskeletal rearrangements in mammalian cells^{6,7}.

PI(3)K activity may be involved in many cellular functions, including mitogenesis, protection from apoptosis, actin-filament organization and protein trafficking. In yeast, the only detectable PI(3)K is Vps34p, which catalyses the conversion of phosphatidylinositol to PtdIns(3)P (ref. 8). Although in mammalian cells there are PI(3)K enzymes that catalyse the formation of phosphatidylinositol-3,4-bisphosphate and phosphatidylinositol-3,4,5-trisphosphate⁹, PtdIns(3)P is the most abundant PI(3)K product in these cells too².

We have shown previously that the binding of EEA1 to endosomal membranes depends on the presence of PtdIns(3)P (ref. 1). To determine whether a direct interaction occurs between EEA1 and PtdIns(3)P, we measured the binding of fusion proteins of EEA1 and glutathione S-transferase (GST) to liposomes. A fusion protein containing the carboxy-terminal

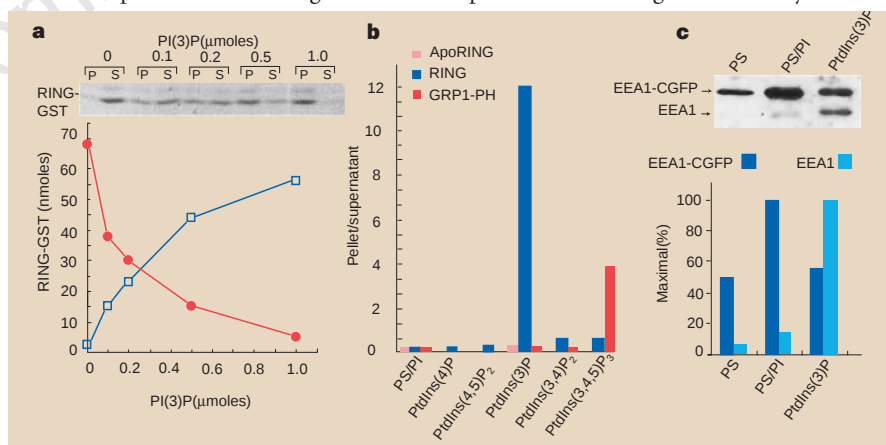


Figure 1 Liposome-binding assays demonstrate the specificity and affinity of EEA1-PtdIns(3) interaction. **a**, Binding of a fusion protein containing the carboxy terminus of EEA1 to liposomes. The RING-GST fusion protein (amino acids 1277–1411 of EEA1 constitute the RING domain) was generated using amplification by PCR of the full-length complementary DNA. Liposomes were composed of phosphatidylserine and phosphatidylinositol (PS/PI) and PtdIns(3)P (PI(3)P) (0–1% of the total lipid, at the expense of phosphatidylinositol, which decreased from 50% to 49%). Liposomes (100 μ moles total lipid) were incubated for 15 min at room temperature with 3 μ g RING-GST in a total volume of 150 μ l (ref. 1). After centrifugation, liposome pellets (P) and supernatants (S) were analysed by SDS-polyacrylamide gel electrophoresis and Coomassie blue staining (inset). Values shown are derived from densitometry of stained bands. **b**, Binding of EEA1 to PtdIns(3)P is specific and requires zinc. A GST fusion protein containing the GRP1 pleckstrin homology domain was generated as described¹⁰. ApoRING was generated by incubation of RING-GST with 50 molar equivalents of methylmethane thiosulphonate, followed by dialysis. Liposomes were composed of PS/PI, and extra phosphoinositides constituted 0.5% of the total lipid at the expense of phosphatidylinositol. Plotted are the ratios of GST fusion protein found in the pellets and supernatants after centrifugation. **c**, A GFP-EAA1 chimera is not specific for PtdIns(3)P. An oligonucleotide encoding GFP was inserted into a unique *NotI* site created in the full-length complementary DNA for EEA1. Cytosolic extracts¹ from Cos-1 cells transfected with EEA1-GFP were treated with wortmannin to inhibit endogenous PI(3)K, and incubated with liposomes as in **b**. Pellets were analysed by immunoblotting with an anti-EEA1 antiserum. Bands corresponding to endogenous, wild-type EEA1 and to the GFP-EAA1 chimera are indicated. The intensity of the bands was plotted.

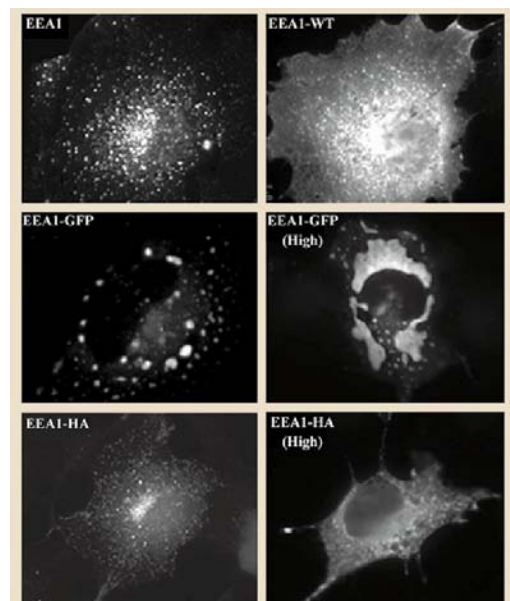


Figure 2 EEA1-GFP prevents the correct formation of early endosomes. Cos-1 cells (control is shown at top left) were transfected with wild-type EEA1 (top right), EEA1-GFP (middle panels) or EEA1-haemagglutinin (bottom panels). Cells were stained with anti-EEA1 antiserum (top panels) or with anti-haemagglutinin antiserum (bottom panels). EEA1-GFP intrinsic fluorescence was visualized in unstained cells (middle panels). Examples of cells expressing relatively low (left panels) or high (right panels) levels of each construct are shown.

135 amino acids of EEA1 bound increasingly to liposomes containing increasing amounts of PtdIns(3)P (Fig. 1a). Chelation of zinc inhibited binding, indicating that the interaction between the fusion protein and PtdIns(3)P was through a RING zinc-finger domain present in this carboxy-terminal region (Fig. 1b). We saw no specific binding to liposomes containing other polyphosphoinositides (Fig. 1b).

A fusion protein containing GST and the pleckstrin-homology domain of GRP-1, which binds phosphatidylinositol-3,4,5-trisphosphate with high affinity ($K_d < 1 \mu\text{M}$) (ref. 10), bound to liposomes containing phosphatidylinositol-3,4,5-trisphosphate, but not to liposomes containing PtdIns(3)P (Fig. 1b). Thus, the binding of the EEA1 RING finger to PtdIns(3)P is specific and of high affinity. Furthermore, the homologous domain in Hrs has been reported¹¹ to bind specifically to PtdIns(3)P. Thus this 'FYVE' domain seems to constitute a conserved binding motif for PtdIns(3)P.

A chimaera containing green fluorescence protein (GFP) at the carboxy terminus of EEA1 (EEA1-GFP) bound to liposomes containing PtdIns(3)P, but, in contrast to wild-type EEA1, this chimaera also bound to liposomes composed of phosphatidylserine or phosphatidylserine and phosphatidylinositol (Fig. 1c).

When transfected into Cos-1 cells, EEA1-GFP caused the formation of lagoon-like structures (Fig. 2), which contained transferrin internalized from outside the cells (not shown), indicating that the structures may represent fused endosomes. Overexpression of wild-type EEA1 (Fig. 2, top), or of another EEA1 chimaera containing nine amino acids of the haemagglutinin protein of influenza virus, did not produce this phenotype. Thus, the loss in binding specificity caused by GFP correlates with the formation of abnormal endosomal

structures, indicating that the specific interaction between EEA1 and PtdIns(3)P may be vital in the control of endosome fusion.

Varsha Patki, Deirdre C. Lawe, Silvia Corvera

Program in Molecular Medicine and
Department of Cell Biology,
University of Massachusetts Medical School,
Worcester, Massachusetts 01605, USA
e-mail: silvia.corvera@ummed.edu

Joseph V. Virbasius, Anil Chawla
Biochemistry and Molecular Biology,
University of Massachusetts Medical School,
Worcester, Massachusetts 01605, USA

1. Patki, V. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 7326–7330 (1997).
2. Auger, K. R. *et al.* *Cell* **57**, 167–175 (1989).
3. Stenmark, H. *et al.* *J. Biol. Chem.* **271**, 24048–24054 (1996).
4. Weisman, L. S. & Wickner, W. J. *Biol. Chem.* **267**, 618–623 (1992).
5. Piper, R. C., Cooper, A. A., Yang, H. & Stevens, T. H. *J. Cell. Biol.* **131**, 603–617 (1995).
6. Komada, M., Masaki, R., Yamamoto, A. & Kitamura, N. *J. Biol. Chem.* **272**, 20538–20544 (1997).
7. Zheng, Y. *et al.* *J. Biol. Chem.* **271**, 33169–33172 (1996).
8. Schu, P. V. *et al.* *Science* **260**, 88–91 (1993).
9. Vanhaesebroeck, B. *et al.* *Trends Biochem. Sci.* **22**, 267–272 (1997).
10. Klarlund, J. K. *et al.* *Science* **275**, 1927–1930 (1997).
11. Gaullier, J.-M. *et al.* *Nature* **394**, 432–433 (1998).

Does practice shape the brain?

Pantev *et al.* (*Nature* **392**, 811–814; 1998) suggest that the degree of cortical reorganization and enhancement of the cortical response to musical notes depends on the age at which musicians first begin learning to play an instrument. Specifically, the younger the subjects were when they started to play, the larger was their cortical reorganization in recognition of piano tones. In addition to its biological interest, such a finding, if true, would have great implications for musical

education. But we believe that the evidence presented by Pantev *et al.* is equally consistent with other interpretations.

Pantev *et al.* based their report on suggested correlations between the age at which musicians first began to play an instrument and the strength of cortical activation in response to piano tones when tested as adults. But the evidence for such correlations in their data is weak.

They present three correlations (for musicians with absolute pitch or with relative pitch, and for both groups combined). The (marginally significant) probability levels quoted are for one-tailed tests, used because 'negative correlations were predicted'. But it is possible that the cortical response to sensory experience might initially increase with age in young children, and so a positive correlation would be predicted. As both negative and positive correlations are possible, two-tailed tests should have been used in this case. Two-tailed tests show none of the reported correlations to be statistically significant at the traditional 5% level, and the confidence intervals for the correlation coefficients all include zero.

Even if more extensive experiments were to find a significant correlation, correlation does not indicate causation. The main contributors to the trends (if they exist) shown in Pantev *et al.*'s Fig. 2 are the musicians who began learning their instrument by 3–5 years of age. Musical tuition in such very young children is almost certainly associated with a high degree of parental musical skill or interest. At least two other explanations for the correlation would therefore be possible. First, perhaps only children with a particular type of cortical response to musical sounds are capable of learning an instrument from a very early age. Such a cortical response may have an inherited component. Second, it is possible that children brought up in musical families hear more music at a very young age, inducing more cortical reorganization. Neither explanation requires the children to have practised an instrument.

Pantev *et al.*'s comment that there was a "dependence of auditory [cortical]... representations on practice beginning before the age of about 10 years" ignores genotypic and environmental influences. It will be possible to establish the true nature of any relationship only by controlling for these effects. In the meantime, given our current state of knowledge, those wanting musical children might be well advised to examine carefully the musical abilities and compact-disc collections of potential mates, rather than investing in expensive music lessons for reluctant three-year-olds.

Pat Monaghan, Neil B. Metcalfe, Graeme D. Ruxton

Institute of Biomedical and Life Sciences,
Graham Kerr Building, Glasgow University,
Glasgow G12 8QQ, UK

Rocky horror picture shows

Deep Impact

A Mimi Leder film

Dreamworks/Paramount: 1998

Armageddon

A Michael Bay film

Touchstone Films: 1998

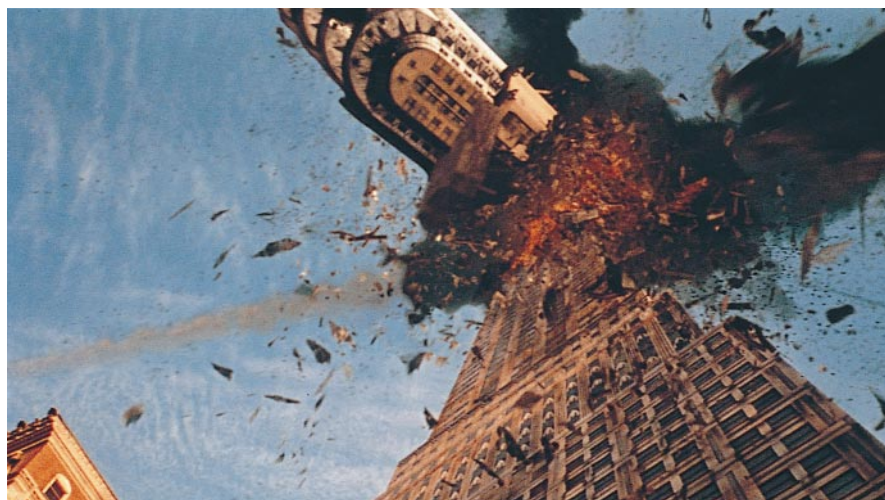
Kevin Zahnle

The penultimate summer of the dying millennium brings us two oddly similar Hollywood disaster movies in which the Earth is threatened by heavenly bodies cast loose from the vault of fixed stars.

In *Deep Impact*, the antagonist is a comet, 10 kilometres in diameter, discovered as a telescopic smudge two years before a predictable and adverse encounter with Earth. The plot centres on the story of humanity saving itself. In *Armageddon*, the antagonist is an asteroid "the size of Texas", discovered 18 days before it is set to annihilate life on Earth. The plot is the story of humanity being saved by Hollywood action heroes. *Armageddon* is probably the more enjoyable film, because it is fast, loud and silly, whereas *Deep Impact* tells its story poorly. But which of them has the better science? *Deep Impact* gets close enough to make it at least possible to discuss these matters. *Armageddon's* fantasy is so complete that any resemblance to the real Universe is accidental. Neither film allows the plot to suffer at the altar of physics.

In *Deep Impact*, astronauts armed with a dozen five-megatonne nuclear devices set off to deflect or disperse the comet a year before impact. As the gravitational binding energy of the comet is less than one megatonne, this is plausibly enough firepower to break it up. The spaceship lands, the astronauts drill holes 100 feet deep in the comet's surface and several nukes are deposited. (The environment at the surface of the comet is not entirely unrealistic. The surface should be blacker, the sky brighter and the nightside lighting diffuse, with the comet's shadow towering like an infinite cylinder into a bright sky of coma and tail. And geysers would probably not shoot off immediately on first light. But overall, it's a credible job.)

Oddly, the astronauts detonate the nuclear devices while the spacecraft is still near the comet, presumably to cripple the craft as a plot device. The shallow explosions do not have the effect one might expect: a thin layer blown off at high velocity, possibly followed by the obliteration of a hemisphere. Instead, the film-makers split the comet into two pieces: a small piece one or two kilometres across for hitting the Earth and generating enormous and spectacular special effects, and a much bigger piece from which the world can still be saved, thereby delivering the happy ending. And all this comes to pass.



Coming down to Earth: science is the first casualty in both *Armageddon* (above) and *Deep Impact* (below).



I have several caveats about the science, of which two stand out. First, the film leaves both fragments on collision course with the Earth; this is wrong. If the original comet was set to collide with Earth, then after the break-up its centre of mass, rather than either of its two fragments, would hit us. Second, the last-minute heroics of the crew would be too late: the big fragment would not blow up harmlessly. There just isn't sufficient time for the pieces to disperse; conservation of energy tells us that the impactor's footprint would widen from 10 to 200 kilometres. The environmental devastation would be comparable to, if not greater than, that from the undispersed comet.

Armageddon's science is simply silly. A few quickies: (1) only the three largest asteroids can be described as "the size of Texas"; (2) at 18 days before impact, a Texas-sized asteroid would be as bright as the stars of Orion's belt, yet somehow it evades discovery until then; (3) the energy required to split the Texas-sized asteroid is 10^{10} megatonnes, roughly a million world nuclear arsenals; and (4) an 800-foot drill-hole (everything in *Armageddon* is bigger) hardly seems like much compared with the vastness of Texas. On the other hand, the meteorite impact that takes out Paris is satisfyingly done.

Armageddon features an almost parodic adherence to the established conventions of Hollywood thrillers. But its attitude towards science (and common sense) is a kind of

ludic nihilism. Meteor showers target major cities for bombing runs. Fashion models live on deep-sea drilling platforms. Space shuttles are launched in pairs and fly in tandem like *Star Wars* fighter planes through swarms of hyper-caffeinated asteroids. Boulders lift off and fly about like flocks of pigeons. Wile E. Coyote dusts off the ashes and opens another box of Acme rockets.

At its heart, *Deep Impact* is a film that trusts in an orderly Universe, in a world of good people ruled by a just god. The movie takes the threat of wayward comets seriously, but it also takes seriously the duty to maintain order in the heavens, and it expects divine intervention on the side of good. By contrast, *Armageddon* is nervous. It inflates the asteroid preposterously, and surrounds it with ridiculous characters and surrealistic action. It is a slapstick comedy based on the oldest joke of all: the humour of pretending that the impossible is real. We make jokes of our fears, and what *Armageddon* fears is a Universe that cares no more for man and his movies than it does for a grain of sand, a Universe that would wipe us out without a thought, a Universe whose lawless laws we will not accept. (Or, on the other hand, maybe they were just trying to make a lot of money.)

Both films betray a touching belief in the power of the government to keep a big secret. But the heavens are no longer a vault of secret mysteries. The stars are not owned by governments and priests, and their courses are no longer decrypted by astrologers alone. Anyone may look, anyone may see, and anyone may compute. This is Galileo's and Newton's message to us all. A comet with Earth's name on it would be known to the world, and followed by the world, from the week it was discovered until the day we died. □

Kevin Zahnle is at the NASA Ames Research Center, Moffett Field, California 94035, USA.

ILL

TOUCHSTONE PICTURES/JERRY BRUCKHEIMER, INC.

The weeds in Fibonacci's garden

Life's Other Secret: The New Mathematics of the Living World

by Ian Stewart

Penguin/Wiley: 1998. Pp. 285. £20, \$24.95

Dennis Bray

This is an engaging, eclectic and strangely disturbing book. At one level it offers a pop-science account of biological patterns — wide-ranging, easy to read and very accessible, even for non-specialists. But beneath this, a subtext repeats insistently, page after page, like the beating of a drum. The message concerns the place of mathematics in our perception of the world, and expresses an almost mystical faith in the marriage of mathematics and living things. We must pursue the mathematical aspects of life, it says, in order to discover “the deep structure of life — what it *really* is”.

Eventually biology will lead us to “a new kind of mathematical theory out there in the intellectual darkness”. Our pursuit will take us to “a grand unified theory of life”, one that “illuminates life on Earth by placing it in a far broader and far less parochial context”. There are echoes here of an idea as old as science itself — the Pythagorean concept of number as the basis of a harmonious Universe.

The pragmatic view that biological forms are subject to physical principles was elegantly illustrated many years ago by D'Arcy Wentworth Thompson in his book *On Growth and Form*. This scholarly discourse on the forms of cells and organisms, the shapes of shells and bones, the arrangements of leaves and petals, appeared in two editions written in 1917 and 1940 and was subsequently abridged by Bonner in 1971. Written in the leisurely style of an earlier generation of academics, Thompson's book is replete with footnotes, quotations in Greek, Latin and French, and pages of numerical data, line drawings and algebraic equations.

Thompson was an incorrigible speculator about the origins of the patterns he saw, and some of his proposed mechanisms have not lasted well. His suggestion that the patterns seen in mitosis are due to electrical force fields, for example, seems quaint today, now that we know about the structure of chromatin and the mitotic spindle. It's hardly surprising that someone writing before the discovery of DNA and without any real appreciation of proteins would get some things wrong, you might say — unless you happen to think that mathematics can explain everything.

Ian Stewart uses *On Growth and Form* as a springboard for his book, but is soon

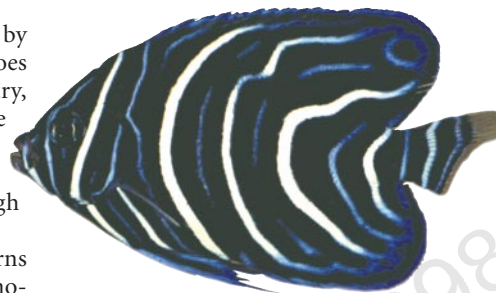
carried into realms undreamt of by Thompson. Mathematically, Stewart goes way beyond simple algebra and geometry, into chaos, complexity and catastrophe theory, cellular automata, neural nets, genetic algorithms, symmetry breaking and more — although oddly enough without giving a single equation.

He sees mathematical patterns everywhere: in the folding of a haemoglobin molecule and the nucleation of a mitotic spindle, in the swarming of ants and the flocking of birds, in the skin markings of leopards and fish, the growth of forests, the aggregation of slime moulds, the swirling colours of a Belousov-Zhabotinsky reaction, and the jewel-like simulations of artificial life.

Everywhere, living and would-be living systems display mathematical rules that express, he feels, the essence of the interplay between genes and the physical Universe. The two great secrets of life, he declares, are DNA and biomathematics. If you understand genes and know the equations, you have life in your grasp.

Well, er, what about all the rest? Where do we fit imaginal disks and photosynthesis, excretion and anaphylactic shock? What about symbiosis, sex and the seventh cranial nerve? Where, in short, are all the diverse, profoundly non-mathematical attributes that we recognize as life? Stewart never confronts this question squarely, but here and there one can glean answers. One reply is that the messy, non-mathematical biological features are just irrelevances. One can ignore them because they in fact contribute nothing to the deep mathematical understanding of the organism, which is our ultimate aim.

The second, slightly more forceful, answer is that chemistry, anatomy, packing of cells, facial expressions and so on are themselves intrinsically mathematical. Although they seem to us untidy and



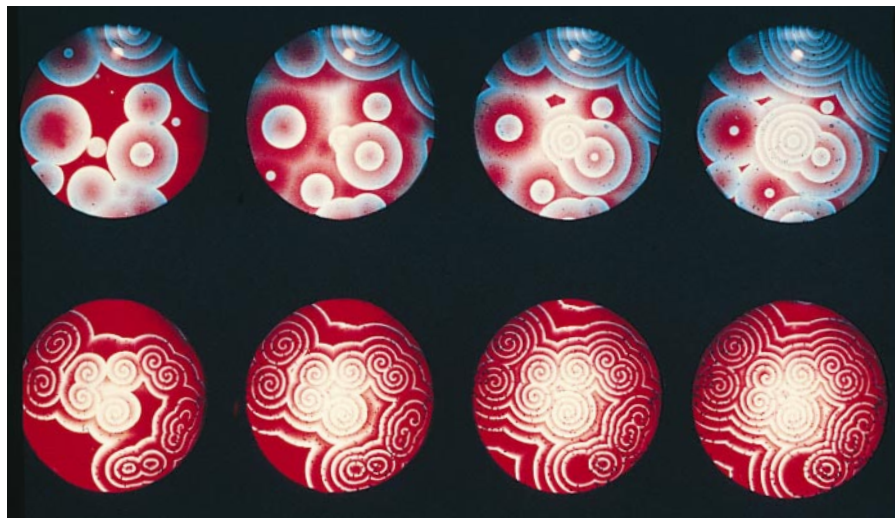
Angelfish: the pattern of stripes changes in accordance with simple mathematical equations.

intractable, there will come a time, in the distant future, when everything will be made plain. The living world will become “just one more page in the Universe's catalog of patterns”.

Examined closely, this argument is quite insidious. It must be true that at some level living organisms can be described in mathematical terms: everything is made of atoms, so in principle one should be able to write down a set of equations governing the development of a fruitfly or the aggressive behaviour of a cichlid fish. But how useful would such an exercise be? Maybe the inhabitants of Zarathustra and Apollonbetnees II (Stewart's fantasy planets) find juggling millions of wave equations intellectually satisfying. But here on planet Earth we do things differently.

Stewart writes — it must be said — remarkably well. It is easy to be carried along by his enthusiasm and seductive prose. But he makes so many dramatic statements and arresting claims, and works on such a broad canvas, that it is natural to wonder how seriously we should take him. His account of the machinery of cell movements is, in my opinion, inaccurate and distorted — but perhaps he is stronger in areas I know less well?

A more general concern lies in the patterns themselves. Human observers have a well-documented propensity to idealize



The Belousov-Zhabotinsky reaction: in a shallow dish containing the correct mixture of chemicals, rings spontaneously form and radiate, mutually annihilating when they collide.

ARTHUR WINFREE/SPL

form and to suppress or ignore small deviations from symmetry.

Twenty years ago a flurry of papers in prestigious journals announced that, when cells in tissue culture divide, the shapes and migration tracks of the daughters display parallel or mirror symmetries. These reports were fascinating because they raised the possibility of an intrinsic pre-programming of cell shape and migration. But unfortunately they were also wrong. Later analysis showed that the same degree of symmetry was claimed by observers looking at tracks on a computer screen — even when these were generated by a completely random process. It would be interesting to see a similarly rigorous analysis of some of the patterns reported in this book. The fearful tiger's face may not be as symmetrical as we suppose, and the much-vaunted Fibonacci series of leaves and petals may in fact be quite difficult to find in the average weed patch. I am not saying that regularities do not exist — just that these rules are not obeyed with great precision. Can they be that important to the organism?

Living organisms occupy space and time, and possess volume and mass. They are subject to the same laws of physics and chemistry as snowflakes, salt crystals and other physical objects, and like these physical objects are expected to display regularities and patterns. Indeed, the complexity of living systems, their capacity for unlimited division and tendency to homeostasis, should make them especially prone to geometrical or mathematical regularities.

But, far from being a “secret of life”, these patterns are what living things have in common with the rest of the Universe. They exist by default and are immediately abrogated if there is any selective advantage to be gained — as in animal mimicry by flowers, for example.

Fascinating, intriguing, beautiful, mathematically pleasing though these patterns are, they are of limited use in understanding the machinery of life. In most cases they simply reveal the operation of some well-established physical principle, such as the close packing of spheres, or surface tension, or the diffusion of chemicals. Useful to know, but hardly a key to unlock further mysteries.

Where would we be if, back in 1917, biologists had read that the forms of dividing cells were governed by the same simple equations as soap bubbles and promptly abandoned all further interest in the matter? We'd know a lot less about the molecular basis of cell division and cell shape — although Pythagoras would have approved! □

Dennis Bray is in the Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

Bouffi à la girafe

Zarafa: A Giraffe's True Story from Deep in Africa to the Heart of Paris

by Michael Allin

Walker: 1998. Pp. 224. \$22. To be published in the UK by Headline in October, £12.99

Philippe Janvier

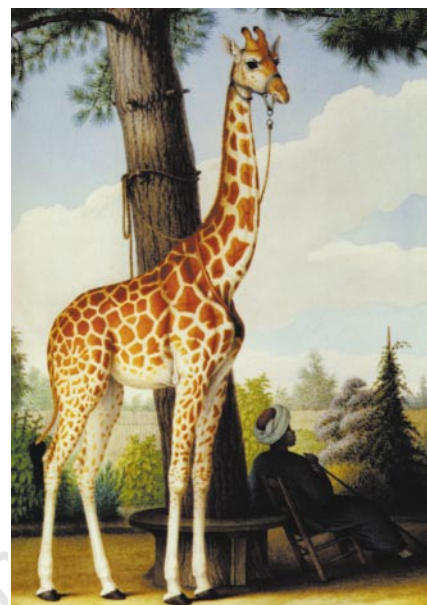
The expression “It's either that or comb the giraffe...” is still used in France to refer to a boring job one has to do. I would certainly not apply it to the task of reviewing this book, which is an admirably documented and charming account of the long journey made by the first giraffe to arrive alive in France.

From Sudan to Paris, the giraffe's progress was accompanied by considerable public interest. This was in 1827, during the reign of Charles X, who had been given the giraffe by the viceroy of Egypt, Muhammad Ali. During her long walk from Marseille to Paris, the giraffe, whom Allin calls Zarafa (the Arabic for giraffe, from a word meaning ‘lovely one’), attracted crowds of people. She finally arrived safely at the Jardin des Plantes, where she lived until 1845. For several years, the giraffe was at the heart of Parisian life. Fashions under ‘à la girafe’ — men's cravats knotted under the chin and women's hair piled up so high that they had to sit on the floors of their carriages. Zarafa is now stuffed and resides in the natural history museum at La Rochelle, in western France.

Allin places Zarafa's story in the political and scientific context of the time: Napoleon's adventurous conquest — and defeat — in Egypt, the rise of Egyptology, the restoration and ultimate fall of the monarchy in France, and the intense scientific activity in the Muséum National d'Histoire Naturelle in Paris. Among the prominent characters are the French consul general in Egypt, the famous tomb robber Bernardino Drovetti, who organized Zarafa's journey to Marseille, and Etienne Geoffroy Saint-Hilaire, a renowned professor at the Muséum National d'Histoire Naturelle.

Geoffroy Saint-Hilaire started his career at the museum at the age of 21, and in his first year there, in 1793, he was faced with the problem of accommodating in the Jardin des Plantes a number of privately owned exotic animals that a revolutionary law had ‘freed from slavery’. He solved this problem by starting the first public menagerie, which at first he had to maintain with his own money. Although he was 55 and a world-famous scientist when the giraffe arrived in Marseille, Geoffroy Saint-Hilaire again took personal charge, following the precious animal step by step, for 21 days, to Paris. As those who have experienced French bureaucracy will know, this do-it-yourself attitude was certainly the best way of ensuring Zarafa's survival.

Allin should perhaps have speculated



Zarafa with her lifelong Sudanese attendant, Atir — both great hits with the ladies of Paris.

about the reasons for Zarafa's fame; after all, a living giraffe that arrived in London the same year went almost unnoticed. One reason for the giraffe-mania in Paris may have been the vivid interest of the French at that time in natural history, prompted by the writings of Buffon, which remain popular to this day (see *Nature* 390, 37; 1997).

Moreover, the neck of the giraffe remained the symbol of Lamarck's ideas on evolution, which were dear to the heart of some learned French people, including Geoffroy Saint-Hilaire.

This charming and never boring account reads like a novel. I recommend it to historians of science and — why not? — to scientists, especially those at the top of their careers, who will learn from Geoffroy Saint-Hilaire how to take matters in hand. □

Philippe Janvier is at the Laboratoire de Paléontologie, Muséum National d'Histoire Naturelle, 8 Rue Buffon, 75005 Paris, France.

The is and the ought

The Ethics of Science: An Introduction

by David B. Resnik

Routledge: 1998. Pp. 221. £45, \$75 (hbk); £14.99, \$24.99 (pbk)

Brian Martin

Scientists know they should not fabricate data or plagiarize the work of others. But what about gift authorship for the head of the lab, making exaggerated claims about results on a grant application, or keeping results secret for reasons of commercial confidentiality? There are no ethics committees to address such topics.

Ethical issues abound in scientific work,

but few scientists have formally studied ethics. They develop most of their responses to ethical issues by observing how specific problems are dealt with in practice, leaving no room for ethical theory.

Yet this *ad hoc* way of learning about ethics leaves much to be desired. What does a scientist learn when supervisors and colleagues use sloppy lab techniques, divert grant funds for undeclared purposes, appoint cronies or lovers over more talented applicants, provide hostile referee reports on competitors' papers, or fail to report some findings that might damage the sales of a profitable drug?

One way of promoting more ethical behaviour in science is to get science students to study ethical issues: David Resnik's book is an excellent text for this purpose. After an overview of ethical theory and the nature of science as a profession, it tackles many ethical issues, including those involving error, conflict of interest, allocating credit, harassment, recruitment, research on humans and animals, and secrecy.

Resnik adopts a traditional idea of science as a quest for objective knowledge, in which honesty and openness are central. His basic model is drawn from academic science, which poses some difficult dilemmas for researchers working in industrial and military settings.

He deals with secrecy in industrial and military science but not with some wider issues, such as whether it is ethical to devote enormous resources to achieve a slight improvement in sales potential or weapons efficiency when millions of people are suffering from lack of attention to basic needs. However, his framework can be used to look at such issues.

Resnik is a philosopher: his main concern is not what scientists do, but what they should do. Practising scientists may therefore be impatient with his analysis, which makes few concessions to the practicalities of surviving and succeeding as a researcher in a climate in which misrepresentation, bias and exploitation are commonplace.

How many scientists, after all, can afford to be entirely truthful about all their findings (including mistakes), share all their data, never give an undeserved citation (for example, to a supervisor or possible referee), support all whistleblowers and avoid any conflict of interests due to funding? Resnik's book perhaps needs to be supplemented by a practical manual on how to survive as an ethical scientist.

He provides a very balanced treatment of the issues, but is sometimes so careful as to be rather dry — despite the spice of topics such as the Baltimore affair and cold fusion. One limitation, though, is that most of the examples and references he cites come from the United States.

David Resnik is right to say that change

will not occur until more scientists behave ethically, and that raising ethical issues with students is a good place to start. Saving the best for last, the book's appendix contains 50 case studies that should provoke many stimulating discussions. □

Brian Martin is in *Science and Technology Studies*, University of Wollongong, NSW 2522, Australia.

The FAQs of astronomy

Just Visiting This Planet: Merlin Answers More Questions about Everything under the Sun, Moon, and Stars

by Neil de Grasse Tyson

Doubleday: 1998. Pp. 336. \$12.95 (pbk)

The Astronomy Cafe: 365 Questions and Answers from "Ask the Astronomer"

by Sten Odenwald

W. H. Freeman: 1998. Pp. 252. \$14.95, £12.95 (pbk)

Jay M. Pasachoff

From the planet Omnisia, Neil de Grasse Tyson's modern-day Merlin answers questions from all comers, or at least from those who subscribe to *Star Date*, a magazine produced by the McDonald Observatory of the University of Texas.

Tyson, a Princeton astrophysicist who also directs the Hayden Planetarium of the American Museum of Natural History in New York, has developed his Merlin persona over the past 15 years. As he supervises the construction of the new planetarium in New York, the old one having met the wrecker's ball, his influence in teaching astronomy to the general public, including children, is as great as anyone's in the United States.

Merlin's answers are short, accurate, sometimes cute and a bit arch. Most of them are straightforward. "Dear Merlin, Couldn't the fragments of comet Shoemaker–Levy 9, when they exploded on the surface of Jupiter back in 1994, cause the massive planet to move?" is answered by "An elephant is a massive animal. Jupiter is a massive planet. A gnat that flew full speed ahead and slammed broadside into an elephant would be about one million times more effective at shoving the elephant than comet Shoemaker–Levy 9 was at shoving Jupiter."

Merlin, who tells us that he has come from the Andromeda galaxy, sometimes takes us to meet an old friend of his, such as Edwin Hubble. "Merlin: Ed, what should curious people know about your recent work? Hubble: I used the excellent hundred-inch Mount Wilson reflecting telescope here to measure the distances to 18 [galaxies]..."

Merlin serves a variety of useful purposes. Sometimes he debunks pseudoscience; he

allows his friend Daniel Fahrenheit to explain his temperature scale (some four dozen such friends are described); and he tells us that the total solar eclipse of 11 August 1999 will be visible from Europe.

It is a pleasure to be able to find answers, from such an authoritative source, to what computer people now call FAQs (frequently asked questions). I had only a few quibbles with Merlin's answers. His explanation of the green flash which is sometimes visible at sunset is incomplete. And he makes the common mistake of referring to the site of the Palomar Observatory as Mount Palomar, instead of Palomar Mountain; there is no Mount Palomar.

Merlin is a good companion for students of any age — from 8 to 80, as the circus used to put it. His earlier book, *Merlin's Tour of the Universe* (Columbia University Press, 1997), has still more questions and answers.

Sten Odenwald, an astronomer specializing in infrared observations, takes questions from all comers at his website, the Astronomy Cafe (www2.ari.net/home/odenwald/cafe.html). Thousands of people have used the site's "Ask the Astronomer" feature; in the book, Odenwald provides answers to 365 of the questions. His answers are clear and accurate, and less flippant in tone than Merlin's; young children apparently do not visit the Astronomy Cafe. It is a wonderful site for non-specialists, and I am glad that the US space agency NASA has helped to support it for a while.

I even learned a few things from Odenwald's book, notably that I have substantially less than five billion years of life on Earth to look forward to. Odenwald explains how expected changes in the Sun's output may lead to the extinction of life here in only 300 million years or so.

Odenwald's book has a few diagrams scattered throughout, and a central eight-page colour section with a nice assortment of pictures. But two typographical errors in the captions are two too many.

Most of Odenwald's answers deal with topics in astronomical science. But one chapter gives answers to practical questions, such as where to get ideas for science-fair projects, or information about a particular artificial satellite. The penultimate chapter has more philosophical replies to questions such as what astronomy will be like in the twenty-first century. The final chapter is about careers in astronomy, and can be recommended to any high-school student. "What is your favourite part of astronomy?", someone asks Odenwald. He replies: "The thrill of the hunt."

We are lucky to have two such interesting books for people to dabble in. Order a dozen of each for birthdays and holidays. □

Jay M. Pasachoff is at the Hopkins Observatory, Williams College, 33 Lab Campus Drive, Williamstown, Massachusetts 01267, USA.

Gaia and natural selection

Timothy M. Lenton

Evidence indicates that the Earth self-regulates at a state that is tolerated by life, but why should the organisms that leave the most descendants be the ones that contribute to regulating their planetary environment? The evolving Gaia theory focuses on the feedback mechanisms, stemming from naturally selected traits of organisms, that could generate such self-regulation.

Organisms alter their material environment and their environment constrains and naturally selects organisms. This connection indicates feedback between life and its environment. The Gaia theory¹ proposes that organisms contribute to self-regulating feedback mechanisms that have kept the Earth's surface environment stable and habitable for life. Gaia theory seeks to explain these mechanisms and how they arise. Natural selection², acting on faithful replication of inherited variation, determines that the organisms that dominate are the ones that leave the most descendants. Together, natural selection and Gaia pose a puzzle: how can self-regulation at the planetary level emerge from natural selection at the individual level³?

Here I attempt to address this question by focusing on the feedbacks to biospheric growth and selective pressures that can arise from environment-altering traits of the biota. This approach helps to bridge the spatial and temporal gaps between the operation of natural selection and the mechanisms of planetary regulation. The arguments are largely restricted to climate regulation, but the principles may be more generally applicable to any environmental variable affecting growth or selective pressures. An example of a feedback on the growth of organisms is the biological amplification of rock weathering, which enhances the uptake of atmospheric carbon dioxide and planetary cooling. Such a feedback tends to stabilize habitable conditions, but can respond only passively to external forcing. To illustrate how biospheric feedbacks affect natural selection, and to explore the impact of random mutation, I extend the Daisyworld model that has underpinned much of the early modelling work on the Gaia hypothesis. In Daisyworld, natural selection can both generate and contribute to environmental regulation. The daisies alter their environment in the same way at the individual and the global level. Therefore, the traits selected at the individual level are ones that change the global environment in a manner favourable to growth.

Moving from models to reality, I discuss how organisms may alter their environment to their benefit, using land ecosystems including rainforest, boreal forest and peat bog as examples. I consider both externally triggered and internally driven vegetation–environment feedbacks. Next, I illustrate the continuing challenge of linking an individual trait to its global consequences by examining the factors affecting dimethyl sulphide production by marine phytoplankton. I discuss the limitations of existing models and make proposals for further testing Gaia theory. At this stage, it seems that natural selection can form an integral part of planetary self-regulation and, where destabilizing effects arise, they may be less likely than stabilizing effects to attain global significance or persist.

Origin of the Gaia hypothesis

In attempting to find a physical basis for detecting the presence of life on a planet, Lovelock⁴ recognized that most organisms shift their physical environment away from equilibrium. In particular, organisms use the atmosphere to supply resources and as a repository for waste products. In contrast, the atmosphere of a planet without life (forced only by solar ultraviolet radiation) should show less disequilibrium (attributable to photochemical processes). Hence, the

presence of abundant life on a planet may be detectable by atmospheric analysis. The atmospheres of Mars and Venus are dominated by carbon dioxide and are only in a mild state of disequilibrium^{5,6}. In contrast, the atmosphere of the Earth is in an extreme state of disequilibrium in which highly reactive gases, such as methane and oxygen, exist together at levels that are different by many orders of magnitude from photochemical steady states⁶ (Fig. 1a). Large, biogenic fluxes of gases are involved in maintaining such disequilibrium (Fig. 1b). This perturbed state is remarkable in that the atmospheric composition is fairly stable over periods of time that are much longer than the residence times of the constituent gases, indicating that life may regulate the composition of the Earth's atmosphere⁷. This concept became the foundation of Gaia theory⁷.

Present conditions at the surface of the Earth are within the relatively narrow boundaries that eukaryotic, multicellular organisms can tolerate⁷. For example, of the major atmospheric gases, nitrogen maintains much of the atmospheric pressure and dilutes

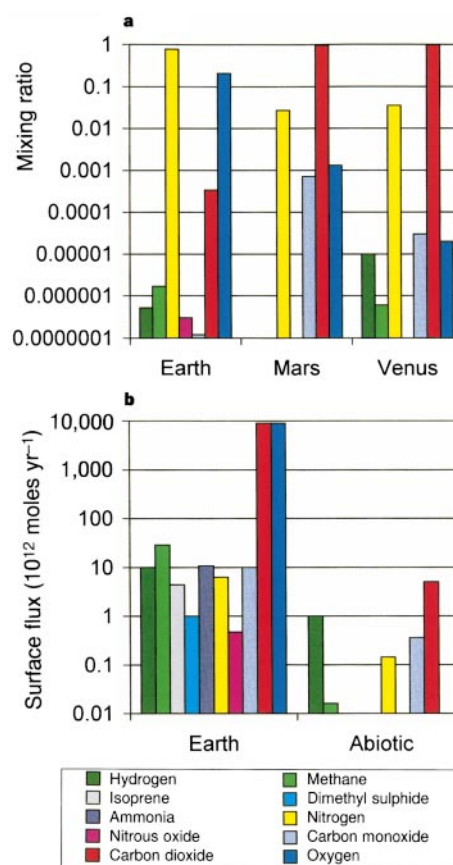


Figure 1 The effect of life on the Earth's atmosphere. **a**, Atmospheric compositions⁹³ of Earth, Mars and Venus (excluding water vapour and noble gases). **b**, Estimated fluxes of gases at the Earth's surface in teramoles (10¹² moles) per year, with (pre-industrial) life^{77,94,95} and without life⁹⁶.

oxygen, which constitutes 21% of the atmosphere, just below the fraction at which fires would disrupt land life⁸. However, oxygen is sufficiently abundant to support the metabolism of large aerobes, and the continuous fossil charcoal record indicates that the oxygen mixing ratio has varied little over Phanerozoic time^{1,8}. The gaseous nitrogen reservoir is largely maintained by the actions of denitrifying organisms, whereas abundant oxygen is the product of past photosynthesis.

Life on Earth has persisted despite major changes in solar forcing⁷. The Sun is thought to have warmed by about 25% (ref. 9) since the origin of life on Earth, over 3.8 billion years ago¹⁰. Assuming a constant radiative response of the Earth to this increased solar input, the Earth's surface temperature would be predicted to have risen by 18 °C, yet the present average temperature is only 15 °C. The continuous habitability of the Earth in the face of a warming Sun indicates that life may have been involved in regulating the climate (although a purely geochemical mechanism is also possible¹¹).

Lovelock and Margulis^{12–14} therefore proposed the Gaia hypothesis of “atmospheric homeostasis by and for the biosphere”, adding that both the redox potential and the acidity of the Earth's surface are anomalous, compared with our planetary neighbours, and can be tolerated by life. The hypothesis included regulation of both atmospheric composition and climate, and suggested roles for the major biogenic gases. The Gaia hypothesis was used to make predictions, for example that marine organisms would make volatile compounds that can transfer essential elements from the ocean to the land. The discovery that dimethyl sulphide¹⁵ and methyl iodide¹⁶ are the major atmospheric carriers of the sulphur and iodine cycles, respectively, supported this suggestion. Later, the Gaia hypothesis was extended¹⁷ to include regulation of much of the chemical composition of the ocean¹⁸.

Development of the Gaia theory

Most criticisms of Gaia have focused on the need for evolutionary mechanisms by which regulatory feedback loops could have arisen or be maintained. One criticism was that the hypothesis implies teleology, some conscious foresight or planning by the biota³. Another was that the Earth is not a unit of selection, and therefore Gaian properties cannot be ‘adaptations’ in a strict neo-Darwinian sense as they cannot be refined by natural selection¹⁹. The challenge is then to explain how Gaian properties could arise from natural selection at lower levels.

The Daisyworld model^{20,21} (Box 1) provided a hypothetical example of planetary regulation emerging from competition and natural selection at the level of individuals, and showed that self-regulation does not necessarily imply teleology. Daisyworld also offered the beginnings of a mathematical framework for understanding self-regulation. The evolved Gaia theory²² recognized that self-regulation is a property of the whole system of life tightly coupled to its environment, and replaced earlier suggestions, which included apparent teleology, that regulation is ‘by and for the biota’^{23,24}. Gaia theory aims to be consistent with evolutionary biology and views the evolution of organisms and their material environment as so closely coupled that they form a single, indivisible, process²⁵. Organisms possess environment-altering traits because the benefit that these traits confer (to the fitness of the organisms) outweighs the cost in energy to the individual.

Since the genesis of the Gaia hypothesis, predictions, proposed mechanisms and evidence consistent with self-regulation have accumulated^{1,24,26}. In particular, the continuous habitability of the Earth, despite major, periodic perturbations, is consistent with the existence of a planetary self-regulating system¹. Planetary impacts²⁷ and volcanic outbursts²⁸ (which may be triggered by the impacts²⁹) appear to have caused mass extinctions³⁰ and climate change³¹ and yet, in all cases, diverse, widespread life and a tolerable climate returned within a short period of geological time.

Are there alternative explanations for what appears to be evidence of self-regulation? Perhaps the antithesis of Gaia is that it is random chance that life has persisted on Earth. A long history of life on Earth may not be evidence in itself for self-regulation: it is merely a prerequisite for conscious observers to have evolved³². However, given the large perturbations and changes in forcing of the Earth's surface, being ‘just lucky’ is a less probable explanation for the persistence of life than the existence of some form of planetary self-regulation.

An intermediate position accepts the existence of regulatory mechanisms on Earth but denies that there is any evolutionary tendency towards planetary self-regulation^{3,33}. In this case, we inhabit a planet that just happened to have stabilizing feedbacks. Where destabilizing feedbacks dominate, life is likely to perish before evolution produces conscious observers. This is consistent with the view that life and climate evolve together^{33,34}, with life adapting to environmental changes, some of which it creates. If an inherent tendency towards self-regulation can be shown it will provide a more probable explanation of the persistence of life on Earth than that of co-evolution.

Some geochemists have asserted that there is no need to invoke life to explain the maintenance of habitable conditions on Earth³⁵. For example, they argue that abiotic, purely geochemical and geophysical feedbacks (Fig. 2) are enough to maintain a favourable climate, offering silicate-weathering negative feedback¹¹ as an

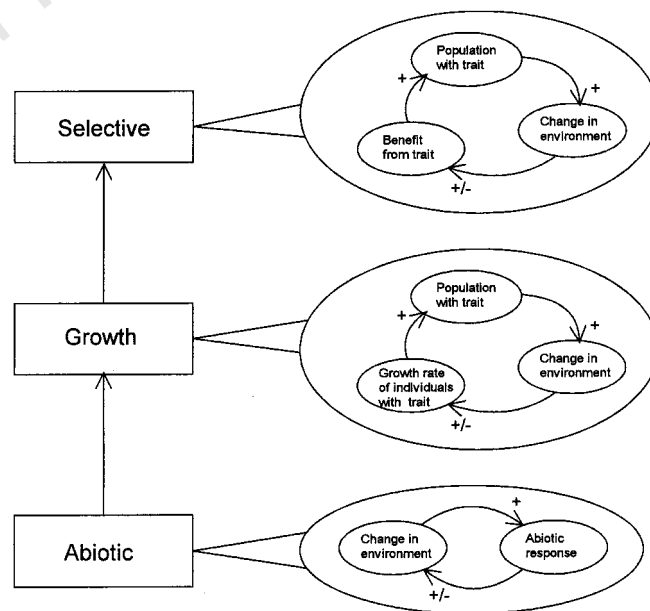


Figure 2 A hierarchy of environmental feedbacks. Three levels are identified, abiotic (purely geochemical and geophysical) feedback, feedback on growth and feedback on natural selection (see text). ‘Trait’ always refers to an environment-altering trait, and ‘growth’ includes reproduction. At each level, both positive and negative feedbacks are possible, and these are illustrated in a general form. A plus symbol indicates a direct relationship. For example, an increase in the population with a particular environment-altering trait increases the resulting change in the environment. A plus/minus symbol indicates a relationship that can be either direct or inverse depending on specific conditions. For example, a change in the environment may increase or decrease the growth rate of individuals carrying the responsible trait, depending on the state of the environment and the direction in which it is being altered. When all the links of a complete feedback loop are positive, the feedback is positive; when one link is negative, the feedback is negative. Steps up the hierarchy are often additive. The activities of organisms can alter an underlying geochemical or geophysical feedback, while feedbacks on selection may be superimposed on underlying feedbacks on growth. (The spread of a trait is often subject to a direct positive feedback that is not shown—the larger a population, the larger its rate of growth.)

explanation. The chemical weathering of calcium-silicate rocks by weakly acidic rain liberates calcium ions that may combine with carbon dioxide as solid calcium carbonate. A rise in planetary temperature (for example) would provoke a rise in weathering rate and a counteracting decline in the concentration of carbon dioxide in the atmosphere (until the weathering sink again matched the source flux of carbon dioxide from volcanoes). Negative feedbacks tend to counteract change and thus stabilize a range of conditions. Silicate-weathering should act to maintain liquid water at a planet's surface. However, geochemical negative feedbacks operate slowly and are not very responsive to perturbation.

Positive feedbacks, notably the ice–albedo feedback³⁶, would also be present on an Earth without life. As polar ice-sheets descend in latitude, the resulting increase in planetary albedo (reflectivity) causes cooling, encouraging the ice to spread further. Beyond a certain latitude there is runaway feedback, and the planet becomes completely covered with ice. Such abiotic, positive feedbacks operate over a narrow range and tend to force a system to extreme states. An ice-covered Earth would be uninhabitable for most organisms. The silicate-weathering negative feedback could not prevent runaway of the ice–albedo positive feedback as the latter operates much faster³⁷, indicating that abiotic feedbacks may not be sensitive enough to maintain a habitable climate on Earth for 3.8 billion years.

There are limitations to a purely geochemical view of Earth's climate history. The absence of siderite from palaeosols of over

2.2 Gyr age indicates that levels of atmospheric carbon dioxide in the late Archaean era may have been insufficient to compensate for lower solar luminosity³⁸. Biogenic methane may then have contributed significantly to the atmospheric greenhouse effect^{1,38,39}. Precambrian glaciations appear to have been rare, although glaciation may have occurred at equatorial latitudes at ~ 0.7 Gyr and ~ 2.2 Gyr ago⁴⁰. Recovery of a habitable climate from suggested 'snowball Earth' conditions would indicate a remarkable resilience to perturbation⁴⁰. More evidence is needed to test specific hypotheses for the roles of life in maintaining a habitable climate and in recovery from perturbation. However, it is clear that organisms are involved in many environmental feedbacks on Earth (including silicate-weathering; Box 2), and their effects need to be considered.

The basis of environmental regulation

Gaia theory focuses on three intrinsic and one extrinsic property of living organisms¹. First, all organisms alter their environment by taking in free energy and excreting high-entropy waste products in order to maintain a low internal entropy⁴¹. Second, organisms grow and multiply, potentially exponentially, providing an intrinsic positive feedback to life (the more life there is, the more life it can beget). Third, for each environmental variable, there is a level or range at which growth of a particular organism is maximum. Thus, carbon-based chemistry and the structures of cells, with their lipid membranes, limit the tolerable range of climate and chemistry⁴².

Box 1 Daisyworld: self-regulation without teleology

Daisyworld^{20,21} is an imaginary grey world orbiting, at a similar distance to the Earth, a star, like our Sun, which gets warmer with time. The world is seeded with two types of life, black and white daisies. These share the same optimum temperature for growth, 22.5°C, and limits to growth of 5°C and 40°C.

Initial conditions on the planet are so cold that daisy seeds cannot germinate. As solar forcing increases and the temperature reaches 5°C, the first seeds germinate. The paleness of the white daisies means that they are cooler than their surroundings, hindering their own growth. The

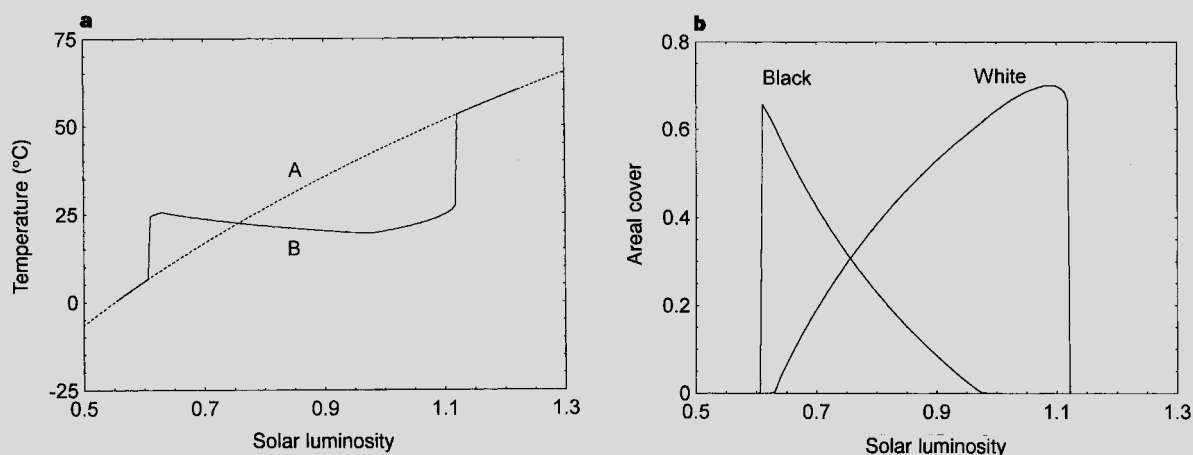
black daisies, in contrast, warm their surroundings, enhancing their growth and reproduction. Hence black daisies come to dominate the initial community (see figure).

As they spread, the black daisies begin to warm the planet. This increases the growth rate of all daisies, an environmental positive feedback that reinforces the spread of life. As the warmer, darker daisies are closer to the optimum temperature than the white daisies, they remain dominant. Soon the limited area of planet surface constrains the explosion of life. When daisies fill the world, the average temperature has risen close to the

optimum for daisy growth.

As the sun warms, the temperature rises to the point at which white daisies begin to appear in the daisy community. As it warms further the white daisies gain the selective advantage over the black daisies and gradually take over. Eventually, only white daisies are left, and when the solar forcing gets too high, self-regulation collapses.

The self-regulation of Daisyworld is impressive: although the solar input changes over a range equivalent to 45°C the surface of the planet is maintained within a few degrees of the optimum temperature for daisy growth.



Box 1 Figure The original Daisyworld model. A thought experiment to show that planetary self-regulation can emerge from natural selection, at the individual level, of types of life with different environment-altering traits^{20,21}. In this case the traits are 'darkness' (albedo 0.15) and 'paleness' (albedo 0.65) of black and white daisies, on a grey planet (albedo 0.4). The equations are

described in ref. 21. **a**, Planetary temperature as solar luminosity increases. The dashed line (A) shows the temperature in the absence of daisies and the solid line (B) shows the temperature in the presence of daisies. **b**, Areal cover of black and white daisies.

Finally, once a planet contains different types of life (phenotypes) with faithfully replicated, heritable variation (different genotypes) growing and competing for resources, natural selection determines that the types of life that leave the most descendants come to dominate their environment².

There is considerable variation in the degree of environmental alteration, rate of growth, environmental constraints and operation of natural selection on different types of life⁴². The argument I develop here focuses on common features of all organisms. Organisms both alter and are constrained by their environment, so feedback is inevitable. Feedback begins at the individual level, but growth implies that feedback has the potential to spread to the global level; the disequilibrium of the Earth's atmosphere indicates that this has occurred.

Gaia theory treats the ways in which organisms alter their environment as (extended¹⁹) phenotypic traits, although, in most cases, the genetic basis of these traits is unknown. From a Gaian perspective, it is the environmental alteration, not which particular organism is responsible for it, that is important. Selection acting on different organisms altering the environment in the same way, to a similar degree, has little environmental consequence.

Some activities that alter the environment are so advantageous (to the organisms carrying out the activities) that they become widespread, fundamental properties of organisms. (An example is photosynthesis, the implications of which have been studied by modelling the Archaean–Proterozoic transition^{25,43}.) Other activities are favourable only under particular environmental conditions and hence are subject to selection. In such cases, it is often changes in one environmental variable that determine whether a trait remains selectively favourable. If the spread of the trait alters this environmental variable, it also alters the forces of selection determining its own value.

A hierarchy of feedbacks. A conceptual hierarchy of feedbacks is used to attempt a synthesis of natural selection and environmental feedback. Abiotic feedbacks, present on a lifeless planet, form the base, added to which are two key steps (Fig. 2). The first step is to add organisms that alter their environment in a manner that affects their growth, without altering the forces of selection on the responsible trait. The resulting ('non-selective') feedbacks on growth rate can be positive (amplifying) or negative (damping). The second step is to consider the spread of traits that alter the forces of selection on themselves. This adds 'selective' feedback (which 'involves selection'). Selective feedback occurs whenever the spread of a trait critically alters the environmental variable that determines the benefit of that trait. The alteration in the variable can maintain, or even promote, conditions in which the trait is advantageous, generating positive selective feedback. Alternatively, the environmental alteration can start to reduce the advantage of the trait, generating negative selective feedback.

Feedbacks on growth. Changes in the environment due to a particular trait can be favourable or unfavourable for growth. However, the spread of a trait that alters the environment in a manner that is favourable to growth tends to be reinforced (positive feedback on growth). In contrast, the spread of a trait that alters the environment away from optimal growth conditions is restrained (through negative feedback on growth). This represents a natural tendency towards environmental self-regulation that can be illustrated by the biological enhancement of silicate weathering^{4,47}. This widespread activity of life on land amplifies the existing geochemical feedback. It is arguably the best established Gaian mechanism and is modelled as a non-selective feedback on growth in Box 2.

To illustrate the first step up the feedback hierarchy (Fig. 2), rock-weathering organisms are introduced to a world in which silicate weathering is the only geochemical feedback (Box 2 Fig.). If the temperature is initially above the optimum for growth (as would be the case for the present Earth without this biological enhancement⁴⁶), the spread of life would increase the rate of rock

weathering and reduce the level of carbon dioxide in the planet's atmosphere, causing cooling. As the temperature decreases and comes closer to the optimum for growth, organisms rapidly fill the world and the temperature drops quickly to below the optimum for growth. Any further spread of life becomes restricted because further cooling reduces growth rate. (If temperatures are initially below the optimum for growth, the spread of rock weathering immediately restricts itself, by reducing growth rates. This is consistent with Gaia theory, because in the real world there are many environment-altering traits. If a warming activity arises under such conditions its spread is reinforced.)

The tendency towards self-regulation is the result of the three intrinsic properties of life. Altering the environment changes existing geochemical and geophysical feedbacks and creates new feedbacks. Growth amplifies any feedback involving organisms. Biological amplification makes geochemical and geophysical feedbacks more sensitive and responsive to perturbation. Environmental constraints determine the form of the function relating growth to each environmental variable. Where there is a range of maximum growth, positive and negative feedback regimes exist on either side of the range. Across a range of external forcing, a system with sufficient biological amplification automatically stabilizes in the negative feedback regime. Hence, in contrast to a dead world, the introduction of organisms brings an inherent tendency to stabilize conditions that are inhabitable by life.

Non-selective feedback on growth provides a basis for environmental self-regulation. However, one such feedback alone can respond only passively to external forcing. As illustrated by the rock-weathering model, the steady state of the system moves in parallel with changes in external forcing, although at a more amenable level.

Feedbacks on selection. The next step up our hierarchy of feedbacks (Fig. 2) is to consider what happens when different types of life with heritable variation arise on a planet and there is competition and natural selection. Daisyworld²¹ (Box 1) provides a simple, albeit hypothetical, example of this situation. The daisy traits of 'darkness' and 'paleness' change the world in a way that alters the forces of selection on them, generating selective feedback.

In Daisyworld, the environmental variable determining selective forces is temperature. In the beginning, when it is cold, being black confers a selective advantage. The spread of the trait of 'darkness' increases the temperature of the world (providing positive feedback on growth). However, this increases the individual temperatures of the daisies, and thus reduces the selective value of 'darkness'. When daisies fill the world the black daisies have brought the planetary temperature to a point where 'darkness' no longer confers much, if any, of a selective advantage. As the sun warms, the white daisies gradually take over.

Daisyworld shows that selective feedback is the result of a trait altering the same environmental variable at the level of selection and at a large scale. In the original Daisyworld, the selective feedbacks are negative. The alteration of the environment resulting from the spread of a trait is one that ultimately reduces the benefit of that trait. This is a consequence of the colour traits altering the environmental conditions in the same direction at the individual and the higher (in this case, global) level.

What happens when the environmental alteration occurs in different directions at the individual and global levels? An example in Daisyworld are black daisies that generate white clouds as a consequence of the convective heat rising from the warm, black daisy clumps²¹. These daisies become warm but, by producing clouds, cool the global environment. As a consequence, the black daisies out-compete the white daisies, driving them to extinction. Thus black cloud-makers maintain the cool conditions in which they are at a selective advantage, an example of positive selective feedback.

Daisyworld shows that natural selection can contribute to enviro-

onmental self-regulation. The combination of selective feedbacks shows active regulation: as the solar forcing increases, the surface temperature of the planet is held close to constant.

Evolution on Daisyworld

Evolution by natural selection needs inherited variation on which to act, and in the original Daisyworld such variation is minimal (there are only two shades of daisy). A Daisyworld with a much larger pool of variant daisies of many different colours, has been modelled⁴⁸. The consequence of competition on Daisyworld is exclusion of all but one or two types of daisy at a given solar luminosity⁴⁹. The types of daisy selected give the planetary temperature most closely matching the optimum for daisy growth. This adds to the case that, given variation, selection can help to generate environmental self-regulation.

Evolutionary biologists argue that there is an inherent conflict between the more immediate, local optimization of evolution by

natural selection and the longer term, larger scale process of environmental regulation⁵⁰. To address this issue properly requires a move from 'static' evolutionary models towards those mimicking the generation of new traits⁵¹.

A specific criticism is that 'cheats' will disrupt self-regulation by not contributing to it and thus saving themselves energy. To address this criticism, Lovelock⁴³ introduced into Daisyworld a grey daisy that saved energy by not producing black or white pigment. This cheat did not destroy regulation because it had a selective advantage only when the solar input was close to the level at which regulation was not required. At extremes of solar input, the cost of producing the right pigment is outweighed by the benefit from being at a temperature closer to the optimum for growth. This emphasizes that environmental regulation can only emerge from traits that are more beneficial than costly to the individual.

A different proposal for local optimization has been made⁵². When the average temperature of Daisyworld is at the optimum for

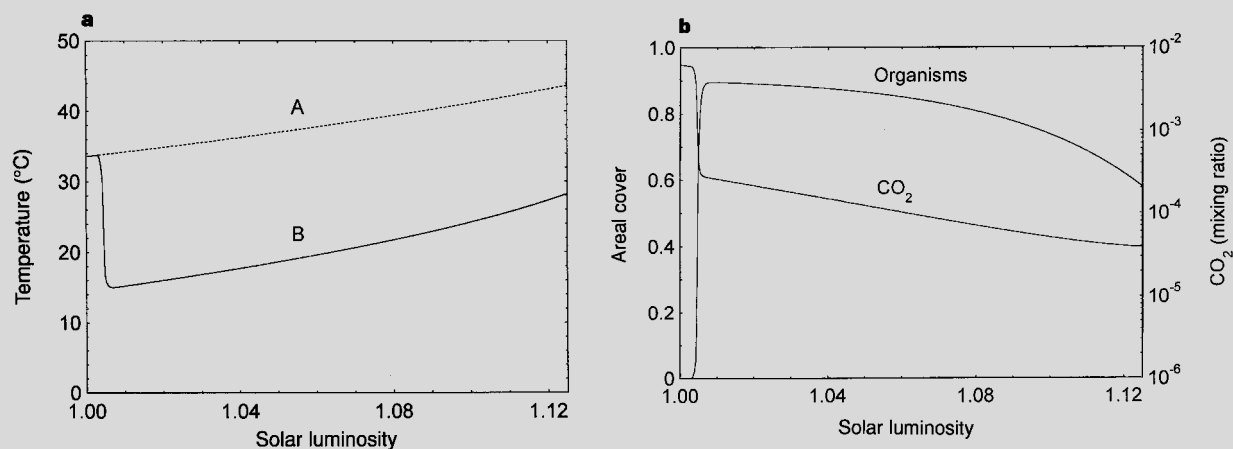
Box 2 Biological amplification of rock weathering

Rock weathering is enhanced by a range of organisms in different ways⁴⁷. There is debate over which effects are most significant. Root respiration and the microbial decay of organic matter enhance soil carbon dioxide levels by 10- to 100-fold over atmospheric levels⁴⁶. The resulting increase in carbonic acid accelerates the weathering reaction. Physical weathering of rocks is enhanced by microbial microfracturing of mineral grains, production of polysaccharides (which swell when wet) by bacteria and fungi and rock splitting by plant roots. These processes create surfaces for chemical weathering. Organic and inorganic acids, actively produced in bacterial

decomposition and by lichen, donate protons which accelerate weathering. The initial formation and subsequent stabilization of soil maintains a high surface area of minerals in contact with acidified water, encouraging further weathering.

From an evolutionary perspective, the biological mechanisms of enhancing rock weathering fall broadly into two categories. Some are the passive result of other processes, whereas others involve an active input of energy. Respiration is part of the fundamental biochemistry of life. The resulting increase in weathering is an inadvertent side effect, which is unlikely to be selected against. In contrast, when microbes and lichens are coloniz-

ing new rock surfaces⁹⁷, several active processes that are subject to selection, such as the production of acids, are used. Where weathering is active, the pursuit of nutrients may be the biological motivation. The global consequence of enhanced rock weathering is a lowering of carbon dioxide levels and hence of temperature. Changes in either of these variables do not alter the selective advantage of seeking nutrients, but they can alter growth rates. Thus, amplification of weathering provides an example of a non-selective feedback on growth (see figure).



Box 2 Figure A simple model of biological amplification of weathering. Rock-weathering land organisms are introduced to a world in which negative feedback on silicate weathering determines the partial pressure of carbon dioxide and the surface temperature. Solar input forces the model, starting at today's value for the Earth and increasing linearly by a total of 12.5%. Carbon dioxide is assumed to be in steady state and the input of carbon dioxide is held constant. Hence, the level of carbon dioxide adjusts so that its removal by weathering matches its input. The radiative relationship between solar luminosity, carbon dioxide level and planetary temperature is a grey atmosphere approximation, fitted to results (L. L. Brown, personal communication) from a one-dimensional, radiative-convective, climate model⁹⁸. The functional dependencies of weathering rate on temperature and carbon dioxide level are those suggested in ref. 11. However, I conservatively⁴⁶

assume that, at present, land organisms amplify weathering rates by a factor of ten. This makes the model world, without life, initially 19°C warmer than today's Earth. Land organisms are introduced after luminosity has increased by 0.2%. The land area covered by organisms is modelled in the same way as in Daisyworld. The weathering rate increases linearly with the area covered by organisms, up to a maximum of ten times the abiotic rate. The optimum temperature for growth is 22.5°C and the relationship between growth rate and temperature is a gaussian curve. Carbon dioxide fertilizes growth. **a**, Planetary temperature evolution, (A) without and (B) with life. **b**, The areal cover of land organisms and the resulting changes in carbon dioxide mixing ratio. (The model is robust to instantaneous perturbations that remove 90% of the organisms. The decline in plant life at the end is due to a lack of carbon dioxide.)

daisy growth, individual black daisies are warmer, and individual white daisies cooler, than their environment. If there is enough heritable variation in optimum temperature among the daisies, we expect the average optimum temperature of the black daisies to evolve upwards, and that of the white daisies to evolve downwards. The gain from this adaptation is a small increase in the area of the planet covered by daisies (at a given solar luminosity in the range of regulation). The loss is a reduction in the range of regulation. In this case, local optimization compromises regulation but does not destroy it.

A problem with all the variations of Daisyworld discussed above is

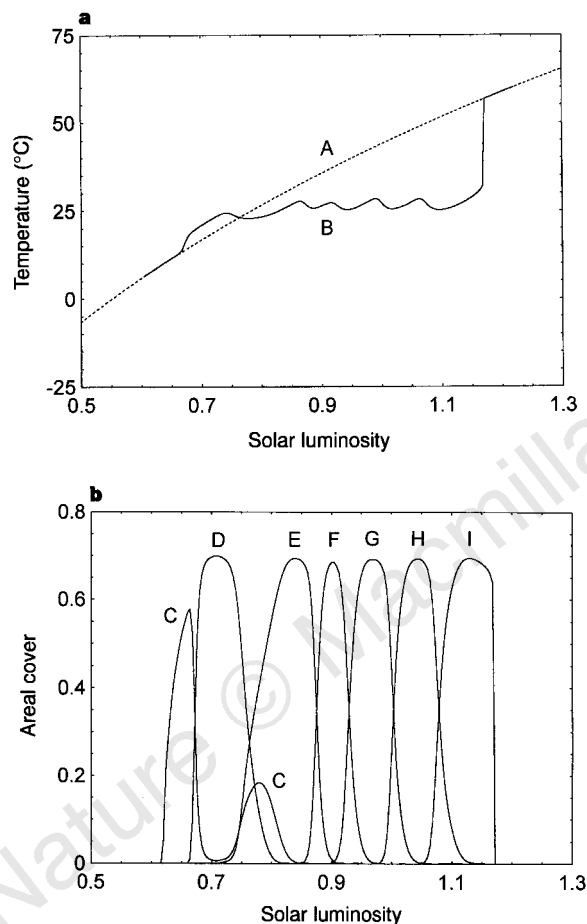


Figure 3 Model for mutating Daisyworld. A model world is seeded with a grey daisy (C) of the same albedo (0.4) as the planet's surface. Daisy albedo can mutate in discrete steps of 0.05 (with limits of 0.2 and 0.65), and the resulting variation is faithfully replicated. The probability of mutation for any individual reproducing daisy is a constant parameter. Thus, the probability of mutation for a particular daisy type is proportional to the population of that type. This means the more abundant types are more likely to beget new types. Mutation is equally probable in either direction of increasing or decreasing albedo. Solar luminosity is increased in steps of 0.004. At each time step there is the possibility of mutation, which is followed by 100 cycles of population dynamics, so the resulting types can approach equilibrium. **a**, Planetary temperature evolution. The dashed line (A) shows the temperature increase without life, and the solid line (B) shows the result of the introduction of daisies (with a mutation probability of one-fifth their areal coverage). **b**, Areal cover of the various daisy types generated. C is the starting type (albedo 0.4) which flourishes initially, when alone, and has a significant population later, when solar luminosity is such that there is no need for regulation. D is a darker mutant of albedo 0.35 that arises and dominates at low solar luminosity. E, F, G, H and I are paler mutants of albedos 0.45, 0.5, 0.55, 0.6 and 0.65, respectively. E–I are favoured in sequence when solar luminosity increases above background temperatures above the optimum for daisy growth.

that the types of daisy are determined at the beginning, with no allowance for further evolutionary change⁵³. To address this criticism⁵¹ I introduce a model of mutation in Daisyworld (Fig. 3). The world is seeded with a grey daisy of the same albedo (reflectivity) as the planet's surface; this daisy makes no contribution to regulation. Colour mutation in either direction, increasing or decreasing albedo, is equally likely. At the individual level there is equal probability that 'Gaian' and 'anti-Gaian' behaviour will arise. Under the initial, cold conditions, a paler daisy would be 'anti-Gaian' in the sense that it would alter the environment away from ideal conditions, whereas a darker daisy would be 'Gaian' in that it would make the environment more favourable to growth. If 'Gaian' and 'anti-Gaian' behaviours are equally likely to emerge at the individual level, what grounds are there to expect self-regulation⁵⁴?

Application of the model shows that with sufficient probability of mutation to generate new types, planetary self-regulation emerges (Fig. 3). A trait that brings the resulting organism closer to the optimum growth conditions will spread. Such a trait is, by definition, 'Gaian'. In contrast, a mutation in an 'anti-Gaian' direction will have its spread restricted by putting the organism responsible at an evolutionary disadvantage. This behaviour results from the daisy traits of albedo altering the environment in the same way at the individual and global levels. More complex studies^{55,56} show that mutation of albedo can also extend the range of regulation.

One can imagine situations in which the outcome of natural selection does not contribute to environmental regulation. An example would be mutation of the temperature tolerance of the daisies⁵². Such internal, physiological adaptations to prevailing conditions may reduce the need to alter the environment. However, there are environmental constraints on such adaptation, such as the onset of water-stress. In this case, there may also be energetic reasons to favour albedo change over physiological change⁵⁵. The outcome of the potential for both environmental alteration and internal adaptation to prevailing conditions deserves further testing. However, there are many examples of living plants altering climate to their own benefit (see 'Land ecosystems' section below).

The selective pressures exerted by different types of organism on one another (for example, predators on their prey) are not directly related to the environment. Will they, therefore, undermine environmental regulation? A developing series of Daisyworld models, with increasingly complex ecology^{48,57,58} (Box 3), provides a framework to address this puzzle. Introducing selective herbivory on the daisies slightly impairs regulation but does not destroy it, because selective pressures from the environment dominate. In contrast, when feedback to the environment is arbitrarily removed, regulation of both population dynamics and climate disappear⁵⁷. In addition, when different types of herbivore are present, operating under different feeding strategies⁵⁸, the dominant herbivore is determined by the daisy–environment feedback, rather than the distribution of daisy types being determined by the herbivores (Box 3).

From models to reality

The robust self-regulation of Daisyworld is an outcome of the direct and strong coupling of plant growth to planetary temperature, but the real world is far more complex. The primary value of models is heuristic⁵⁹. The actual mechanisms of climate regulation in Daisyworld were not intended to explicitly represent those on the Earth, although similar feedbacks were later recognized¹ (see below). However, Daisyworld has been usefully adapted to modelling of community ecology^{48,57,58} (Box 3) and of chemical and climatic regulation^{1,25,43,60}.

Daisyworld ignores important levels of the environment between the individual and the global levels. The temperature of an individual daisy is determined only by the difference between its albedo and the albedo of the planet. If the daisies were real, their individual temperatures would depend somewhat on their neighbours and

their region. In a two-dimensional extension of Daisyworld⁵⁶, albedo mutation generates a heterogeneous distribution of daisy colours and the inclusion of spatial heat transport extends the range of climate regulation.

The simple models presented here show a homogeneous world, in which the background conditions are the same everywhere, but the real world has many regions and a range of background conditions. The clearest regional distinction is between land and ocean. Superimposed on this is the varying solar input across the spherical surface of the Earth; this creates a continuum of background temperature and moisture conditions between the Equator and the poles.

Oversimplifying to a one-niche model means that when the system collapses, it does so in one rapid switch. In the two-dimensional Daisyworld⁵⁶, destructive habitat fragmentation impairs temperature regulation once a threshold is reached at which the diminishing areas of daisies become disconnected. This emphasizes the importance of unrestricted competition and natural selection for regulation on Daisyworld. In reality there are many niches, and physical barriers can prevent organisms from competing directly with one another. By modelling separate land and ocean biotas, it can be shown that the demise of regulation in one region may not spell disaster for the entire system, as long as enough organisms in other areas are contributing to regulation⁶⁰.

Land ecosystems

In the real world, coupling between life and its environment occurs at all scales, beginning with localized niche construction⁶¹ and with ecosystems emerging as integrated systems between the individual and the global levels^{62,63}. Gaian feedback concepts may usefully be applied to these levels. Ecosystem-level environmental feedbacks must be understandable in terms of natural selection. Equally, an ecosystem cannot spread and persist indefinitely if it alters regional and global conditions away from those amenable to the constituent organisms²². Ecosystems that have stabilizing feedback will tend to persist and spread, whereas ecosystems that develop destabilizing feedback will tend to collapse and disappear. Thus, we might expect ecosystems with stabilizing internal and environmental feedbacks to predominate (see 'Testing Gaia' section below).

The trees of the Amazon rainforest, through generating a high level of water cycling, maintain the moist environmental conditions in which they can persist⁶⁴ (a positive feedback on growth and selection). Nutrients are also effectively retained and recycled²². If too much forest is removed, the water-regulation system can collapse, the topsoil is washed away and the region reverts to arid semi-desert, a change that may be difficult to reverse⁶⁵.

Boreal forest trees are somewhat analogous to the dark daisies of Daisyworld¹. The individual trees possess the traits of snow-shedding and darkness that give them a low albedo and make them warmer than their surroundings. The presence of forest warms the region and the hemisphere⁶⁶. However, the forest remains well below the optimum temperature for growth for most of the year. Thus, the system shows constrained, positive feedback (on growth). This amplifies any increases in temperature due to regional warming. The winter mean temperature change for 1965–1995 shows a 2.5°C rise across the band of northern high-latitude forests⁶⁷ and further amplified warming is predicted under global change⁶⁸.

The terrestrial biomes significantly alter climate in different ways^{1,65,66} and compete for space. Thus, their geographical distributions are the result of complex feedbacks involving climate and succession. The resulting dynamic balance can shift in response to both external triggers and internally driven changes.

Shifts in the balance between boreal forest and tundra amplify external forcing: 115,000 years ago, orbital forcing reduced summer temperatures and seems to have triggered the spreading of the arctic tundra southwards to replace boreal forest⁶⁹. The resulting increase in albedo, because of unmasked snow cover, would have added to

regional and planetary cooling and may have generated the onset of glaciation⁶⁹. The positive feedback has probably also operated in the opposite direction: 6,000 years ago, orbital forcing warmed the high latitudes, which would have triggered boreal forests to spread northwards and amplify the initial warming⁷⁰. Indeed, boreal forest may often be involved in maintaining periods of global warmth⁷¹.

Internal changes in ecosystems, involving feedbacks on growth and natural selection, may drive changes in climate^{1,72}. Ecological succession may involve the onset of regulatory feedbacks, including resistance to invasion by damaging outsider species⁵¹. For example, peat bogs have been proposed as a 'climax' ecosystem in many regions⁷². Through promoting soil acidification, iron-capping, water storage and the build up of peat, peat bog plants exclude trees and other plants^{72,73}, generating positive selective feedback.

Marine phytoplankton

The production of dimethyl sulphide (DMS) by marine phytoplankton is an example of an individual trait with global consequences⁷⁴, which illustrates the complex steps linking organisms to their environment^{75–77}. This topical area of research (reviewed in refs 74–77) has been greatly stimulated by Gaian thinking.

Different species of marine phytoplankton produce varying amounts of dimethylsulphoniopropionate (DMSP), the precursor

Box 3 Ecology and biodiversity in Daisyworld

It was proposed⁹⁹ that the mathematics of evolutionary models would be simpler if the evolution of organisms and of their physical environment was considered as a single process. To test this idea, herbivores then carnivores were introduced into a biodiverse Daisyworld model⁴⁸. These predators fed unselectively, eating the same fraction of their prey, regardless of the prey's abundance. The model shows remarkable mathematical stability for many species. Adding the unselective herbivores results in smaller populations of daisies at any given time and, therefore, a small decrease in the range of temperature regulation. Adding the carnivores reduces the herbivore populations, thus increasing the daisy populations and the range of regulation.

The biodiversity model provides a framework in which to explore the implications of interorganism selection for environmental self-regulation. A variation of this model is to compare the effects of an unselective herbivore (type 0) and three different types of selective herbivore (types 1–3) that favour more abundant over less abundant daisies to varying degrees⁵⁷. The frequency-dependent selection leads to exploiter-mediated co-existence of the daisies and differing degrees of daisy biodiversity according to the precise herbivore feeding strategy. One type of selective herbivore (type 3) improves temperature regulation relative to unselective herbivory, whereas the other two (types 1 and 2) slightly impair regulation.

A next step towards modelling a realistic community is to introduce the three types of selective herbivore together⁵⁸ (large types⁵⁷ 1–3). If selection by the herbivores dominates the system, we expect there to be small populations of many daisy colours right up to the extents of regulation, which would therefore be reduced. Instead at extreme solar luminosities there is an innate tendency for a reduction in daisy biodiversity, because only one or two daisy shades (dark ones at the beginning and light ones at the end) are selected by the environment and can provide regulation. The rising populations of these one or two daisy types determine that the dominant herbivore becomes the one that eats the largest proportion of daisies that are at high abundances (type 1). The model thus shows the herbivores being selected by the daisy–environment feedback, rather than the distribution of daisy types being selected by the herbivores. Furthermore, decreasing the number of food-web connections reduces resilience⁵⁸.

of DMS⁷⁶. DMSP is one of a range of compatible solutes, believed to be produced as an osmolyte, that can alleviate salt-stress and prevent freezing^{76,78}. The conversion of DMSP to DMS is catalysed by the enzyme DMSP lyase. This process is enhanced by virus infection and zooplankton grazing^{76,79} and may be adaptive^{79,80}. The main reservoir of DMS is in the ocean, where it is consumed by bacteria and oxidized to dimethylsulphoxide (DMSO)⁷⁶. Air–sea exchange results in a net flux of DMS to the atmosphere⁷⁶ (Fig. 1b). In the atmosphere, DMS is oxidized in a range of oxidation reactions⁸¹. The main pathway generates sulphur dioxide which is further oxidized to sulphate, and which can ultimately contribute to sulphate aerosol formation. (A secondary pathway generates methanesulphonate which can be measured in ice cores and used as a proxy for DMS production in the past^{82,83}.) Sulphate aerosol is a major source of cloud condensation nuclei^{77,84}, which can form cloud droplets that are important scatterers of solar radiation. DMS-derived aerosols can therefore increase cloud albedo⁸⁵ and the consequent return of solar radiation to space, which would be expected to cool the affected region and the planet⁷⁴.

What types of feedback (Fig. 2) are generated by DMS emissions? Any benefits of altering the environment are estimated to be small in comparison to the energy cost of DMSP production⁷⁸; thus, we might expect DMSP production to remain selectively favourable regardless of the environmental alteration. A (non-selective) negative feedback on growth of DMS-emitting phytoplankton and climate was first proposed⁷⁴ whereby a reduction in temperature and light beneath clouds reduces photosynthesis and restricts the spread of DMS producers. Subsequent modelling⁶⁰ elaborated this proposal²⁶, and indicated that the formation of a thermocline at ~10 °C might limit the supply of nutrients to the surface ocean, thus setting an effective optimum for plankton growth. Beneath this temperature lies the originally proposed regime of negative feedback. Above it, however, an increase in temperature may be amplified by a decrease in photosynthetic production⁸⁶, DMS production and cloud reflectivity, generating positive feedback. Evidence that DMS production in the Southern Hemisphere was enhanced during the last ice age^{82,83} indicated that the feedback may then have been negative, but switched to become positive as temperatures rose at glacial termination^{60,75,83}. However, no single relationship between temperature and DMS emissions is consistent with all existing data⁸³. This is not surprising, given the other factors now thought to influence DMS production. For example, it has been proposed⁸⁷ and confirmed^{76,88,89} that when nitrogen is abundant the compatible solute, glycine betaine, is produced instead of DMSP. When nitrogen is limited, organisms switch to making DMSP. Marine organisms are responsible for a sea–air flux of ammonia of similar magnitude to the DMS flux⁹⁰. In the atmosphere, acidic sulphate cloud droplets scavenge the alkaline ammonia, in a process that may be important for particle nucleation⁹¹, and the rain out of these droplets may fuel new photosynthetic production in remote marine areas⁹⁰.

Testing Gaia

The Gaia theory is a valuable hypothesis generator^{1,23–25}. Predictions from Gaia, including ancillary hypotheses of mechanisms, have been tested and corroborated²⁴. However, direct tests of Gaia⁹² are difficult, because of the temporal and spatial scales involved. A challenge for future work will be to develop a range of more realistic and testable models.

One approach would be to test whether adding a realistic environment and random generation of environment-altering traits would enhance the stability of a community ecology model (W. D. Hamilton, personal communication), thus allowing ecosystem-level predictions to be made. For example, the consequences of nitrogen fixation and biological amplification of rock weathering on available nutrient reservoirs and community dynamics could be tested against natural and experimental systems.

Modelling of interacting ecosystems competing for space in a shared planetary environment might test whether those with stabilizing feedbacks come to predominate. A dynamic vegetation component for future global circulation models (GCMs), being developed at present, could be put to this task and offer comparison with the real world (P. M. Cox, personal communication). GCM simulations indicate that vegetation almost always influences climate for its own benefit, by making high latitudes warmer and by increasing continental rainfall everywhere (R. A. Betts, personal communication). However, GCM simulations are restricted to relatively short timescales. A simpler (energy-balance) model of feedbacks between the biosphere and climate would provide a framework with which to explore effects over longer timescales (for example, those involved in glacial–interglacial transitions). In this context, feedbacks that have yet to be incorporated in GCM simulations (such as those involving DMS production) could be quantitatively evaluated and predictions made. In addition, the effects of competition and selection, within the land and ocean realms, could be explored by explicit modelling of different types of terrestrial vegetation and marine phytoplankton with characteristic environment-altering traits.

Conclusions

When asked to explain how planetary self-regulation could have arisen, we are in much the same position as Darwin when asked how the eye could have evolved. We see a complex phenomenon and have only the beginnings of a theory with which to tackle the puzzle. Darwin focused on the exponential growth of organisms, the constraints imposed on them by their environment and the resulting natural selection. The fact that organisms also alter their environment means there is an inevitable feedback connection between the living and non-living. I have tried to describe the forms that such a connection could take. The implications may be far reaching; simple principles suggest that environmental regulation can emerge at levels from the individual to the global. Natural selection is seen as an integral part of Gaia, and Gaia theory also has something to offer evolutionary biology. Gaian models suggest that we must consider the totality of organisms and their material environment to fully understand which traits come to persist and dominate. □

T. M. Lenton is at the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

1. Lovelock, J. E. *The Ages of Gaia* 2nd edn (Oxford Univ. Press, Oxford, 1995).
2. Darwin, C. *The Origin of Species* (John Murray, London, 1959).
3. Doolittle, W. F. Is Nature really motherly? *CoEvol. Quart.* Spring, 58–63 (1981).
4. Lovelock, J. E. A physical basis for life detection experiments. *Nature* **207**, 568–570 (1965).
5. Hitchcock, D. R. & Lovelock, J. E. Life detection by atmospheric analysis. *Icarus* **7**, 149–159 (1967).
6. Lovelock, J. E. Thermodynamics and the recognition of alien biospheres. *Proc. R. Soc. Lond. B* **189**, 167–181 (1975).
7. Lovelock, J. E. Gaia as seen through the atmosphere. *Atmos. Environ.* **6**, 579–580 (1972).
8. Watson, A. J., Lovelock, J. E. & Margulis, L. Methanogenesis, fires and the regulation of atmospheric oxygen. *Biosystems* **10**, 293–298 (1978).
9. Newman, M. J. & Rood, R. T. Implications of solar evolution for the Earth's early atmosphere. *Science* **198**, 1035–1037 (1977).
10. Mojzsis, S. J. *et al.* Evidence for life on Earth before 3,800 million years ago. *Nature* **384**, 55–59 (1996).
11. Walker, J. C. G., Hays, P. B. & Kasting, J. F. A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *J. Geophys. Res.* **86**, 9776–9782 (1981).
12. Lovelock, J. E. & Margulis, L. Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis. *Tellus* **26**, 2–10 (1974).
13. Margulis, L. & Lovelock, J. E. Biological modulation of the Earth's atmosphere. *Icarus* **21**, 471–489 (1974).
14. Lovelock, J. E. & Margulis, L. Homeostatic tendencies of the Earth's atmosphere. *Origins Life* **5**, 93–103 (1974).
15. Lovelock, J. E., Maggs, R. J. & Rasmussen, R. A. Atmospheric dimethyl sulphide and the natural sulphur cycle. *Nature* **237**, 452–453 (1972).
16. Lovelock, J. E., Maggs, R. J. & Wade, R. J. Halogenated hydrocarbons in and over the Atlantic. *Nature* **241**, 194–196 (1973).
17. Lovelock, J. E. *Gaia—A New Look at Life on Earth* (Oxford Univ. Press, 1979).
18. Whitfield, M. The world ocean: mechanism or machination? *Interdisc. Sci. Rev.* **6**, 12–35 (1981).
19. Dawkins, R. *The Extended Phenotype* (Oxford Univ. Press, 1983).
20. Lovelock, J. E. in *Biomimicry and Biological Metal Accumulation* (eds Westbroek, P. & de Jong, E. W.) 15–25 (Reidel, Dordrecht, 1983).
21. Watson, A. J. & Lovelock, J. E. Biological homeostasis of the global environment: the parable of Daisyworld. *Tellus* **35B**, 284–289 (1983).
22. Lovelock, J. E. Geophysiology: a new look at earth science. *Bull. Am. Meteorol. Soc.* **67**, 392–397 (1986).

23. Kerr, R. A. No longer willful, Gaia becomes respectable. *Science* **240**, 393–395 (1988).
24. Lovelock, J. E. Hands up for the Gaia hypothesis. *Nature* **344**, 100–102 (1990).
25. Lovelock, J. E. Geophysiology, the science of Gaia. *Rev. Geophys.* **27**, 215–222 (1989).
26. Kump, L. R. & Lovelock, J. E. in *Future Climates of the World: A Modelling Perspective* (ed. Henderson-Sellers, A.) 537–553 (Elsevier, Oxford, 1995).
27. Alvarez, W. *et al.* Impact theory of mass extinctions and the invertebrate fossil record. *Science* **223**, 1135–1141 (1984).
28. Officer, C. B., Hallam, A., Drake, C. L. & Devine, J. D. Late Cretaceous and paroxysmal Cretaceous/Tertiary extinctions. *Nature* **326**, 143–149 (1987).
29. Rampino, M. R. Impact cratering and flood basalt volcanism. *Nature* **327**, 468 (1987).
30. Raup, D. M. & Sepkoski, J. J. Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503 (1982).
31. Rampino, M. R. & Volk, T. Mass extinctions, atmospheric sulphur and climatic warming at the K/T boundary. *Nature* **332**, 63–65 (1988).
32. Watson, A. J. Gaia. *New Sci. Inside Science* **48**, 1–4 (1991).
33. Ehrlich, P. in *Scientists on Gaia* (eds Schneider, S. H. & Boston, P. J.) 19–22 (MIT, London, 1991).
34. Schneider, S. H. & Londer, R. *The Coevolution of Climate and Life* (Sierra Club, San Francisco, 1984).
35. Holland, H. D. *The Chemical Evolution of the Atmosphere and the Oceans* (Princeton Univ. Press, NJ, 1984).
36. Budzko, M. I. The effect of solar radiation variations on the climate of the Earth. *Tellus* **21**, 611–619 (1969).
37. Caldeira, K. & Kasting, J. F. Susceptibility of the early Earth to irreversible glaciations caused by carbon dioxide clouds. *Nature* **359**, 226–228 (1992).
38. Rye, R., Kuo, P. H. & Holland, H. D. Atmospheric carbon dioxide concentrations before 2.2 billion years ago. *Nature* **378**, 603–605 (1995).
39. Hayes, J. M. in *Early Life on Earth, Nobel Symposium No. 84* (ed. Bengtson, S.) 220–236 (Columbia Univ. Press, New York, 1994).
40. Evans, D. A., Beukes, N. J. & Kirschvink, J. L. Low-latitude glaciation in the Palaeoproterozoic era. *Nature* **386**, 262–266 (1997).
41. Schrodinger, E. *What is Life?* (Cambridge Univ. Press, 1944).
42. DeDuke, C. *Vital Dust* (Basic, New York, 1995).
43. Lovelock, J. E. Geophysiology. *Trans. R. Soc. Edinb. Earth Sci.* **80**, 169–175 (1989).
44. Lovelock, J. E. & Whitfield, M. Life span of the biosphere. *Nature* **296**, 561–563 (1982).
45. Lovelock, J. E. & Watson, A. J. The regulation of carbon dioxide and climate: Gaia or geochemistry. *Planet. Space Sci.* **30**, 795–802 (1982).
46. Schwartzman, D. W. & Volk, T. Biotic enhancement of weathering and the habitability of Earth. *Nature* **340**, 457–460 (1989).
47. Schwartzman, D. W. & Volk, T. Biotic enhancement of weathering and surface temperatures on earth since the origin of life. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **90**, 357–371 (1991).
48. Lovelock, J. E. A numerical model for biodiversity. *Phil. Trans. R. Soc. Lond. B* **338**, 383–391 (1992).
49. Maddock, L. Effects of simple environmental feedback on some population models. *Tellus* **43B**, 331–337 (1991).
50. Maynard Smith, J. & Szathmari, E. *The Major Transitions in Evolution* (Freeman, Oxford, 1995).
51. Hamilton, W. D. Ecology in the large: Gaia and Genghis Khan. *J. Appl. Ecol.* **32**, 451–453 (1995).
52. Saunders, P. T. Evolution without natural selection: further implications of the Daisyworld parable. *J. Theor. Biol.* **166**, 365–373 (1994).
53. Keeling, R. in *Scientists on Gaia* (eds Schneider, S. H. & Boston, P. J.) 118–120 (MIT, London, 1991).
54. Kirchner, J. W. The Gaia hypothesis: can it be tested? *Rev. Geophys.* **27**, 223–235 (1989).
55. Stöcker, S. Regarding mutations in Daisyworld models. *J. Theor. Biol.* **175**, 495–501 (1995).
56. Von Bloh, W., Block, A. & Schellhuber, H. J. Self-stabilization of the biosphere under global change: a tutorial geophysiological approach. *Tellus* **49B**, 249–262 (1997).
57. Harding, S. P. & Lovelock, J. E. Exploiter-mediated coexistence and frequency-dependent selection in a numerical model of biodiversity. *J. Theor. Biol.* **182**, 109–116 (1996).
58. Harding, S. P. The effects of food web complexity on community stability and climate regulation in a geophysiological model. *Tellus* (submitted).
59. Oreskes, N., Shrader-Frechette, K. & Belitz, K. Verification, validation, and confirmation of numerical models in the Earth sciences. *Science* **263**, 641–646 (1994).
60. Lovelock, J. E. & Kump, L. R. Failure of climate regulation in a geophysiological model. *Nature* **369**, 732–734 (1994).
61. Laland, K. N., Odling-Smee, F. J. & Feldman, M. W. The evolutionary consequences of niche construction: a theoretical investigation using two-locus theory. *J. Evol. Biol.* **9**, 293–316 (1996).
62. Patten, B. C. & Odum, E. P. The cybernetic nature of ecosystems. *Am. Nat.* **118**, 886–895 (1981).
63. Chapin, F. S. III *et al.* Biotic control over the functioning of ecosystems. *Science* **277**, 500–504 (1997).
64. Salati, E. in *The Geophysiology of Amazonia: Vegetation and Climate Interactions* (ed. Dickinson, R. E.) 273–296 (Wiley, New York, 1987).
65. Shukla, J., Nobre, C. & Sellers, P. Amazon deforestation and climate change. *Science* **247**, 1322–1325 (1990).
66. Bonan, G. B., Pollard, D. & Thompson, S. L. Effects of boreal forest vegetation on global climate. *Nature* **359**, 716–718 (1992).
67. Hansen, J., Ruedy, R., Sato, M. & Reynolds, R. Global surface air temperature in 1995: return to pre-Pinatubo level. *J. Geophys. Res.* **23**, 1665–1668 (1996).
68. Betts, R. A., Cox, P. M., Lee, S. E. & Woodward, F. I. Contrasting physiological and structural vegetation feedbacks in climate change simulations. *Nature* **387**, 796–799 (1997).
69. Gallimore, R. G. & Kutzbach, J. E. Role of orbitally induced changes in tundra area in the onset of glaciation. *Nature* **381**, 503–505 (1996).
70. Foley, J. A., Kutzbach, J. E., Coe, M. T. & Levis, S. Feedbacks between climate and boreal forests during the Holocene epoch. *Nature* **371**, 52–54 (1994).
71. Otto-Bliesner, B. L. & Upchurch, G. R. Jr Vegetation-induced warming of high-latitude regions during the Late Cretaceous period. *Nature* **385**, 804–807 (1997).
72. Klinger, L. F. in *Scientists on Gaia* (eds Schneider, S. H. & Boston, P. J.) 247–255 (MIT, London, 1991).
73. Hamilton, W. D. Gaia's benefits. *New Sci.* **151**, 62–63 (1996).
74. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326**, 655–661 (1987).
75. Lovelock, J. E. A geophysiologicalist's thoughts on the natural sulphur cycle. *Phil. Trans. R. Soc. Lond. B* **352**, 143–147 (1997).
76. Liss, P. S., Hatton, A. D., Malin, G., Nightingale, P. D. & Turner, S. M. Marine sulphur emissions. *Phil. Trans. R. Soc. Lond. B* **352**, 159–169 (1997).
77. Andreae, M. O. & Crutzen, P. J. Atmospheric aerosols: biogeochemical sources and role in atmospheric chemistry. *Science* **276**, 1052–1058 (1997).
78. Caldeira, K. Evolutionary pressures on planktonic production of atmospheric sulphur. *Nature* **337**, 732–734 (1989).
79. Wolfe, G. V., Steinke, M. & Kirst, G. O. Grazing-activated chemical defence in a unicellular marine alga. *Nature* **387**, 894–897 (1997).
80. Hamilton, W. D. & Lenton, T. M. Spora and Gaia: how microbes fly with their clouds. *Ethol. Ecol. Evol.* **10**, 1–16 (1998).
81. Ravishankara, A. R., Rudich, Y., Talukdar, R. & Barone, S. B. Oxidation of atmospheric reduced sulphur compounds: perspective from laboratory studies. *Phil. Trans. R. Soc. Lond. B* **352**, 171–182 (1997).
82. Legrand, M. *et al.* Ice-core record of oceanic emissions of dimethylsulphide during the last climate cycle. *Nature* **350**, 144–146 (1991).
83. Legrand, M. Ice-core records of atmospheric sulphur. *Phil. Trans. R. Soc. Lond. B* **352**, 241–250 (1997).
84. Ayers, G. P. & Gras, J. L. Seasonal relationship between cloud condensation nuclei and aerosol methanesulphonate in marine air. *Nature* **353**, 834–835 (1991).
85. Falkowski, P. G. *et al.* Natural versus anthropogenic factors affecting low-level cloud albedo over the North Atlantic. *Science* **256**, 1311–1313 (1992).
86. Roemmich, D. & McGowan, J. Climatic warming and the decline of zooplankton in the California current. *Science* **267**, 1324–1326 (1995).
87. Andreae, M. O. in *The Role of Air–Sea Exchange in Geochemical Cycling* (ed. Buat-Menard, P.) 331–362 (Reidel, Dordrecht, 1986).
88. Turner, S. M., Malin, G., Liss, P. S., Harbour, D. S. & Holligan, P. M. The seasonal variation of dimethyl sulphide and dimethylsulphoniopropionate concentrations in nearshore waters. *Limnol. Oceanogr.* **33**, 364–375 (1988).
89. Gage, D. A. *et al.* A new route for synthesis of dimethylsulphoniopropionate in marine algae. *Nature* **387**, 891–894 (1997).
90. Liss, P. S. & Galloway, J. N. in *Interactions of C, N, P and S Biogeochemical Cycles and Global Change* (eds Wollast, R., Mackenzie, F. T. & Chou, L.) 259–281 (Springer, Berlin, 1993).
91. Eisele, F. L. & McMurray, P. H. Recent progress in understanding particle nucleation and growth. *Phil. Trans. R. Soc. Lond. B* **352**, 191–201 (1997).
92. Watson, A. J. in *Gaia in Action, Science of the Living Earth* (ed. Bunyard, P.) 65–74 (Floris, Edinburgh, 1996).
93. Wayne, R. P. *Chemistry of Atmospheres* 2nd edn (Oxford Univ. Press, 1991).
94. Mackenzie, F. T., Ver, L. M., Sabine, C., Lane, M. & Lerman, A. in *Interactions of C, N, P and S Biogeochemical Cycles and Global Change* (eds Wollast, R., Mackenzie, F. T. & Chou, L.) 1–61 (Springer, Berlin, 1993).
95. Brasseur, G. P. & Chatfield, R. B. in *Trace Gas Emission from Plants* (eds Sharkey, T. D., Holland, E. A. & Mooney, H. A.) 1–27 (Academic, San Diego, 1991).
96. Holland, H. D. *The Chemistry of the Atmosphere and Oceans* (Wiley, New York, 1978).
97. Jackson, T. A. & Keller, W. D. A comparative study of the role of lichens and “inorganic” processes in the chemical weathering of recent Hawaiian lava flows. *Am. J. Sci.* **269**, 446–466 (1970).
98. Kasting, J. F., Whitmore, D. P. & Reynolds, R. T. Habitable zones around main sequence stars. *Icarus* **101**, 108–128 (1993).
99. Lotka, A. *Elements of Mathematical Biology* (Dover, New York, 1956).

Acknowledgements. I thank J. E. Lovelock and S. J. Lovelock for inspiration and support; J. E. Lovelock for earlier versions of the rock-weathering and mutating Daisyworld models; J. Maynard Smith and W. D. Hamilton for encouragement; J. R. Lenton, M. Whitfield, A. J. Watson, S. P. Harding, S. M. Turner, P. S. Liss, T. Tyrrell, C. Barlow and T. Volk for comments on the manuscript; UEA for a research studentship; the Gaia Charity for additional funding; and Plymouth Marine Laboratory for providing facilities.

Correspondence should be addressed to the author (e-mail: t.lenton@uea.ac.uk).

An extraordinary cluster of massive stars near the centre of the Milky Way

E. Serabyn*, D. Shupe† & D. F. Figer‡

* Jet Propulsion Laboratory, California Institute of Technology, MS 171-113, 4800 Oak Grove Drive, Pasadena, California 91109, USA

† IPAC, California Institute of Technology 100-22, Pasadena, California 91125, USA

‡ Department of Physics and Astronomy, UCLA, Los Angeles, California 90095, USA

The relative numbers of newborn stars of different masses in a galaxy (the initial mass function) determines whether the galaxy's interstellar gas goes mainly into long-lived low-mass stars, as in the disks of normal spiral galaxies, or into short-lived massive stars, as has been proposed for "starburst" galaxies^{1,2}. The centre of the Milky Way is not a fully-fledged starburst region, but its star formation rate per unit volume of space is nevertheless roughly a thousand times that of the disk^{3,4}. It is, however, very difficult to study the initial mass function near the centre, because the dust in the gas clouds obscures the starlight, and the relatively rare young stars are mixed with much more numerous older stars¹¹. Here we report high-resolution infrared observations of a compact cluster of stars in the central region of our Galaxy. We find approximately 100 young, massive main-sequence stars, several of which seem to be among the most massive in the Galaxy. This cluster may be a weak analogue of the large star clusters in starburst galaxies, which opens the possibility of studying the starburst phenomenon through a local example.

Where best to look for normal main-sequence stars in the crowded Galactic nucleus? Because of the dense concentration of bright giants (which have finished core hydrogen-burning and left the main sequence) in the nucleus, a compact young cluster is the most viable target. Excluding the cluster in our Galaxy's central parsec because of extreme crowding, the most promising target is perhaps the recently discovered "Arches" cluster^{5,6} (G0.121 + 0.017), a compact concentration of a dozen or so bright stars located (in projection) ~25 pc from the Galactic Centre. These stars have been tentatively identified as rare Wolf-Rayet (WR) stars^{5,6}, which represent a late phase in the evolution of massive stars, by which time rapid mass-loss has largely stripped the stars of their outer layers. As the brief (~2–5 × 10⁶ yr; ref. 7) WR phase occurs only in the most massive stars, WR stars are signposts for recent massive-star formation. Slightly less massive stars still burning hydrogen can thus be expected in their vicinity.

Because the Milky Way's nucleus is completely obscured by intervening dust at optical wavelengths, near-infrared (near-IR) images of the Arches cluster were acquired with the Keck I telescope on a night of excellent seeing (0.45" full-width at half-maximum). Images were acquired at three near-IR wavelengths (J(λ = 1.25 μ m), H(λ = 1.66 μ m) and K'(λ = 2.12 μ m)), and combined into the false-colour image shown in Fig. 1. This near-IR image reveals a very rich stellar cluster, with a substantial overdensity above the surrounding 'background' stars. A radius of ~6" (or 0.23 pc for a distance⁸ of 8 kpc) encompasses most of the stars in the dense cluster core, but a slightly larger radius (~9", or 0.35 pc) is needed to include all of the previously known bright WR-like stars^{5,6}, and so this larger radius may better describe the cluster's full extent. As seen in Fig. 1 and the inset to Fig. 2, the stars in the compact cluster core show very similar near-IR colours, and so share a common distance and extinction. In contrast, the surrounding field stars show a much wider colour range (Fig. 1), implying a

range of extinctions and distances. The average near-IR cluster colour (Fig. 2) implies⁹ an average visual extinction of 28.4 mag, fully consistent with a location near the Galactic Centre.

We address the nature of the Arches stars by means of the cluster's near-IR colour-magnitude diagram (Fig. 2) and number counts (Fig. 3). The number counts show a very clear distinction between the populations interior and exterior to a dividing radius of 9": the exterior field stars are consistent in number and brightness with the population of cool giants expected for this location¹⁰, while the interior (cluster) number counts extend to much brighter stars. Given this contrast, and the presence in the brightest few bins in Fig. 3 of young WR-like stars^{5,6} (we note that Of supergiants resemble WR stars at low spectral resolution¹¹, but these would also be young), an old population such as that found in globular clusters is unlikely.

A young cluster is thus indicated, and indeed the near-IR fluxes from the Arches stars, when corrected for distance and extinction, are consistent^{12,13} with those of the most massive (O-type) main-sequence stars (Fig. 3). In addition, a significant tail extends to even

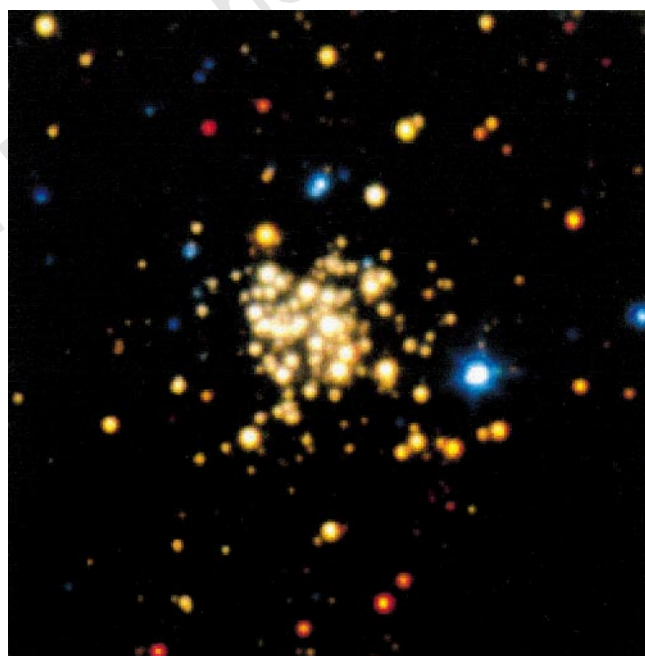


Figure 1 False-colour near-IR image of the Arches cluster constructed from Keck I near-IR camera (NIRC) images acquired on 3 June 1996. Three wavelengths were observed, J (λ = 1.25 μ m), H (λ = 1.66 μ m) and K' (λ = 2.12 μ m). The K' passband²² was used instead of the more standard K band (2.21 μ m) because the former is somewhat narrower, which helps to avoid saturation on the brightest stars. In the image, blue corresponds to J, yellow to H, and K' to red. The field centre is roughly at right ascension 17 h 42 min 39.9 s, declination -28° 48' 13". North is up and west to the right, and the Galactic plane is orientated at ~30° east of north. NIRC has a 256 × 256 detector array with a plate scale of 0.15" per pixel, and a field of view of 38.4 × 38.4". Several slightly offset individual exposures were taken, and the final image also occupies 256 × 256 pixels. Flux calibration was relative to UKIRT faint standard star 34. Data reduction was carried out with the IRAF package, and sky subtraction, flat fielding and the masking of bad pixels were treated in standard fashion. The integration times were ~1 min each at K' and H, and 6 min in the J band which is subject to heavier extinction. The detection limits near the edges of the field were ~21st, 19th and 18th magnitudes at J, H and K', respectively. The blue stars are foreground stars, the very red stars are background stars that have been subject to heavy extinction, while the stars making up the cluster all show a very similar intermediate colour (average $H - K' = 1.43$ mag). (We note that a patch of high extinction occupies the south-western corner of the field.) Stars were counted and fluxes measured with IRAF's DAOPHOT subroutine²³, and 343, 714 and 904 stars were found in the J, H and K' bands.

brighter stars, probably blue supergiants and WR-like stars, which no doubt represent the evolved descendants of the most massive and short-lived of the O stars. This model receives further support from the cluster's colour-magnitude diagram, in which the stars interior to 6" (that is, the stars most clearly part of the cluster) are largely confined to a very well-defined vertical track (Fig. 2). The ten or so brightest stars (those with line emission characteristic of WR-like stars^{3,6}) follow a slightly sloped track, with the brighter stars being slightly redder, while the bulk of the fainter stars below K-band magnitude $K \approx 11$ show a fairly constant colour ($H - K \approx 1.75$) down to $K \approx 14.5$ mag (for $K > 13.5$ mag, the scatter increases due to source confusion), as expected for the upper main sequence^{12,13}. Given the cluster's fairly uniform extinction, and the narrow spread in the observed track on the colour-magnitude diagram (± 0.05 mag), the bluer colours seen at $K > 11$ mag indicate that these fainter stars are probably somewhat hotter than the WR-candidates, consistent with temperatures in the O-star range. We also note that increasing the outer cut-off radius for the colour-magnitude diagram to 9" introduces a known bright, cool giant with CO absorption features⁶ ~ 0.3 mag to the red (right) of the top end of the track in Fig. 2, a location in line with the expected near-IR colour separation between the giant branch and the upper main sequence. Thus, the infrared fluxes and colours of the fainter Arches stars are fully consistent with hot, massive O stars.

Unfortunately, the absolute near-IR fluxes of the brightest main-sequence stars and the associated supergiant and WR phases remain observationally and theoretically somewhat uncertain, due both to their scarcity and their high mass-loss rates. Accurate classifications of individual stars are thus not possible from broad-band fluxes alone. On the other hand, spectroscopic classification can also be ambiguous for some massive stars^{11,14}, making a combination of near-IR fluxes and spectra necessary to definitively classify individual stars. Unresolved multiple stars may also mimic brighter single stars, but the strong dependence of near-IR brightness on stellar mass for hot main-sequence stars¹³ suggests that this cannot skew our results dramatically unless the number of fainter O stars is itself scaled upward by a similarly large multiplicity factor. This would imply many hundreds of fainter O stars, also a rather extreme result.

Thus, even allowing for brightness uncertainties as large as 1 mag, it is clear from Fig. 3 that the stellar population of the Arches cluster consists of an abundance of hot, massive O stars, and that the population extends to the brightest, most massive, and rarest O3 stellar type at the very top of the main sequence, where the distribution cuts off as it makes a transition to more evolved types (O-giants, supergiants and WR stars). For the theoretical fluxes given in Fig. 3, we find the remarkable total of ~ 120 massive O stars (mass > 20 solar masses, $20M_{\odot}$), a dozen of which may have evolved to the WR phase. Even allowing for uncertainties in near-IR fluxes, and in stellar models and ages, $\sim 100 \pm 20$ massive stars remain. This rich cluster thus substantially increases the potential count of exotic, massive stellar types in our Galaxy, making it a prime testing ground for elucidating stellar structure and mass-loss phases near the stellar high-mass instability limit. It is also clear that the relative scarcity of extremely massive stars found to date for our Galaxy is simply the result of extinction to the interesting neighbourhoods, and specifically toward our Galaxy's nucleus.

The total observed mass in O stars in the Arches cluster is then $(5,000 \pm 1,000)M_{\odot}$. Extrapolating to include lower-mass stars by means of a standard power-law initial mass function (IMF) of exponent -2.35 , the total cluster mass becomes $\sim (1.5-6) \times 10^4 M_{\odot}$ for lower mass cut-offs in the range $(2-0.1)M_{\odot}$. Such a mass is comparable to that of a small globular cluster¹⁵. Similarly, the total cluster luminosity is $\sim 10^8$ times the solar luminosity, and the ionizing photon flux a few 10^{51} s^{-1} , making this cluster a powerhouse by Galactic standards. The cluster is also quite young, certainly < 5 Myr if WR stars are indeed present⁷. However, if O3 stars are present and the WR-candidates are instead high-mass-loss O-supergiants, the cluster could be as young as 1 Myr.

A single Galactic cluster containing ~ 100 massive O stars is indeed remarkable, as the number of O stars in the Arches cluster then dwarfs all other known young clusters in the Milky Way's large-scale disk. In particular, both NGC3603, the only giant H II region visible from our perspective in the Galaxy¹⁶, and W49A, probably the most luminous star-forming region in the disk¹⁷ (but which is obscured optically), have O-star tallies of ~ 50 , of which most are of subtypes cooler and less massive than O5 (compare Fig. 3). The

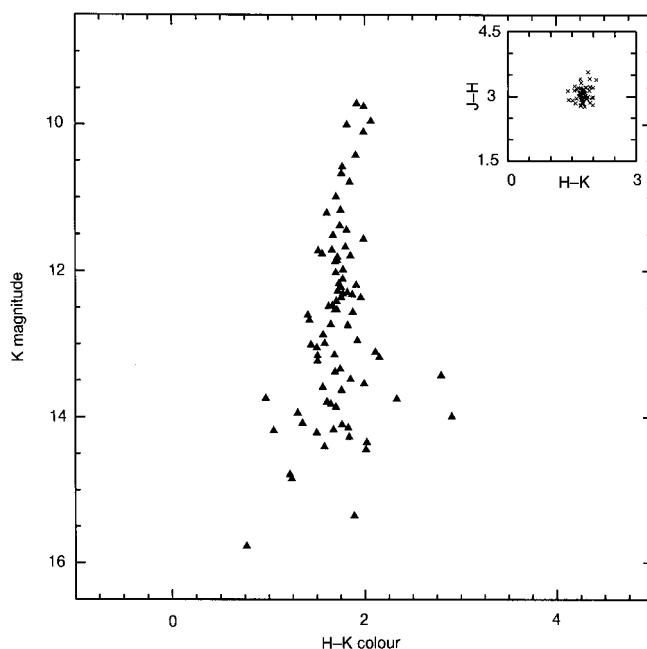


Figure 2 Colour-magnitude diagram for the stars detected at J, H and K' interior to a radius of 6". In both this plot and Fig. 3, K' magnitudes have been converted to K magnitudes using the observed average $H-K'$ colour and the ratio of the H-K and $H-K'$ wavelength intervals²², yielding $K = K' - 0.30$. Inset, infrared colour-colour

diagram for the Arches cluster stars detected at J, H and K'. The stars shown are limited to those within 6" of the cluster centre and $K < 13$ mag, the latter to avoid source confusion effects in the cluster core.

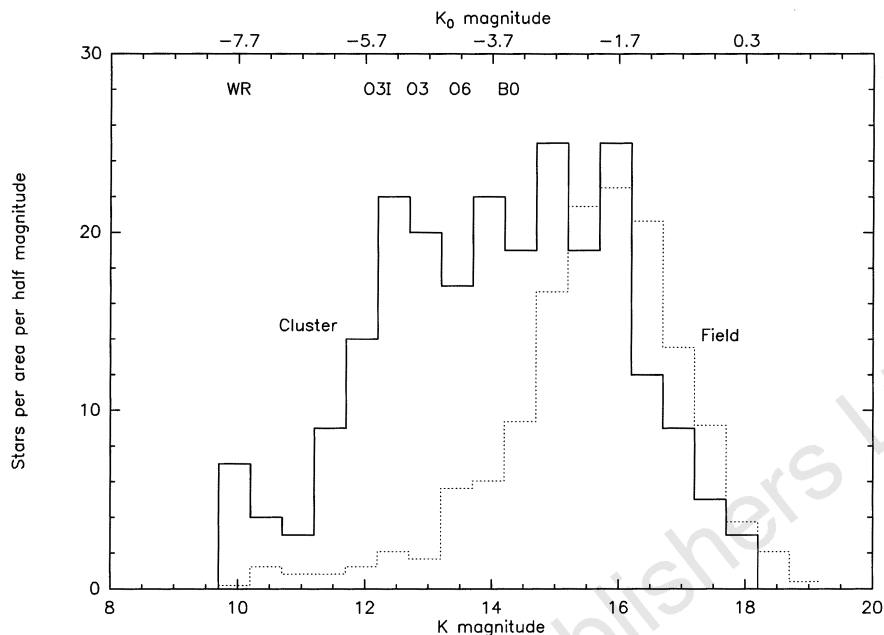


Figure 3 K number counts (per half-magnitude luminosity bin) for the stars inside (solid histogram) and outside (dotted histogram) a 9'' dividing radius. The outer histogram has been rescaled to the same area (81 π arcsec²). The top scale gives the absolute magnitudes, assuming a distance of 8 kpc, and a K-band extinction of 3.2 mag. The main-sequence and supergiant stellar type labels are positioned

assemblage of massive O stars in the Arches cluster is then in fact quite comparable to the recently tabulated O-star population in the rich, young cluster¹⁴ R136 in the spectacular 30 Doradus H II region of the Large Magellanic Cloud. However, the Arches cluster is even more compact: with a radius (~ 0.3 pc) only one-third that of R136 (~ 1 pc), its average stellar density ($\sim 3 \times 10^5 M_{\odot} \text{pc}^{-3}$) exceeds even that in the core of R136. Furthermore, including also neighbouring star-formation regions such as the Sagittarius A, Sgr A East, Sgr B2, and Quintuplet clusters¹⁸, the star formation activity in our Galaxy's obscured nucleus probably bears a close resemblance to the enormous 30 Dor star-formation complex¹⁹.

We now consider whether the IMF in our Galactic nucleus is "normal". The detection of numerous normal O stars accompanying the Arches cluster's previously known WR-like stars clearly removes the need for an overtly bizarre stellar population. However, as our observations do not yet reach stars below $\sim 20 M_{\odot}$, it is premature to reach conclusions regarding the nature of the Arches' IMF, especially as a large number of high-mass stars can arise simply from the rapid production of a large total number of stars with a normal IMF. The most robust result at present is thus the total stellar mass generated in the 'burst', a quantity which is relatively insensitive to the addition of lower-mass stars. One clear implication of a very compact, massive ($\sim 10^4$ – $10^5 M_{\odot}$) young cluster is the need for very efficient conversion of natal molecular-cloud material into new stars, as nearby molecular-cloud clumps²⁰ are comparable in mass. Furthermore, given additional young clusters of comparable total mass located nearby (listed above), the formation of massive clusters of the order of $10^4 M_{\odot}$ is probably the preferred mode of star formation in the nuclear environment. Given the proximity of these clusters to each other, some degree of synchronization or common triggering may be suspected. All of this suggests that nuclear star formation may be triggered by large-scale shock compression of the interstellar medium. If correct, shock-induced star formation may be the unifying mechanism linking young clusters in galactic spiral arms (of total mass 10^2 – $10^3 M_{\odot}$), young clusters in quiescent galactic nuclei (masses $\sim 10^4$ – $10^5 M_{\odot}$), and super star-clusters²¹ in colliding and starburst galaxies (masses

according to refs 12, 13. The WR is arbitrarily located on the brightest bin, as these stars may be of this type^{5,6}. Derivation of an accurate IMF slope is not possible from this data set, as blending in the cluster core remains a problem: identification of stars in restricted magnitude intervals indicates that the stellar counts in the cluster core suffer from incompleteness at K magnitudes as bright as 13.

$\sim 10^5$ – $10^8 M_{\odot}$), the different mass scales then being attributable to the widely different mass accumulation rates in the three environments. Determination of the shape of the Arches cluster IMF at lower masses—which can be pursued at higher resolution both with the Hubble Space Telescope and adaptive optics on large ground-based telescopes—would thus prove an illuminating link in this hierarchy. □

Received 5 January; accepted 7 May 1998.

- Scalo, J. M. in *Starbursts and Galaxy Evolution* (eds Thuan, T. X., Montmerle, T. & Van, J. T. T.) 445–465 (Editions Frontieres, Gif-sur-Yvette, 1987).
- Rieke, G. H., Loken, K., Rieke, M. J. & Tambllyn, P. Starburst modeling of M82: test case for a biased initial mass function. *Astrophys. J.* **412**, 99–110 (1993).
- Güsten, R. in *The Center of the Galaxy* (ed. Morris, M.) 89–106 (Kluwer, Dordrecht, 1989).
- Serabyn, E. & Morris, M. Sustained star formation in the central stellar cluster of the Milky Way. *Nature* **382**, 602–604 (1996).
- Nagata, T., Woodward, C. E., Shure, M. & Kobayashi, N. Object 17: another cluster of emission-line stars near the Galactic Center. *Astron. J.* **109**, 1676–1681 (1995).
- Cotera, A. S. et al. The discovery of hot stars near the Galactic Center thermal radio filaments. *Astrophys. J.* **461**, 750–761 (1996).
- Meynet, G. Wolf-Rayet population syntheses for starburst galaxies. *Astron. Astrophys.* **298**, 767–783 (1995).
- Reid, M. The distance to the center of the Galaxy. *Annu. Rev. Astron. Astrophys.* **31**, 345–372 (1993).
- Rieke, G. H. & Lebofsky, M. J. The interstellar extinction law from 1 to 13 microns. *Astrophys. J.* **288**, 618–621 (1985).
- Becklin, E. E. & Neugebauer, G. Infrared observations of the Galactic Center. *Astrophys. J.* **151**, 145–161 (1968).
- Conti, P. S., Hanson, M. M., Morris, P. W., Willis, A. J. & Fossey, S. J. Of-type stars HD 16691 and HD 190429 show WN-like spectra in infrared K band. *Astrophys. J.* **445**, L35–L38 (1995).
- Wegner, W. Intrinsic colour indices of OB supergiants, giants and dwarfs in the UBVRIJKLM system. *Mon. Not. R. Astron. Soc.* **270**, 229–234 (1994).
- Vacca, W. D., Garmany, C. D. & Shull, J. M. The Lyman-continuum fluxes and stellar parameters of O and early B-type stars. *Astrophys. J.* **460**, 914–931 (1996).
- Massey, P. & Hunter, D. A. Star formation in R136: a cluster of O3 stars revealed by Hubble Space Telescope spectroscopy. *Astrophys. J.* **493**, 180–194 (1998).
- Mandushev, G., Spassova, N. & Staneva, A. Dynamical masses for Galactic globular clusters. *Astron. Astrophys.* **252**, 94–99 (1991).
- Melnick, J., Tapia, M. & Terlevich, R. The Galactic giant HII region NGC 3603. *Astron. Astrophys.* **213**, 89–96 (1989).
- De Pree, C. G., Mehringer, D. M. & Goss, W. M. Multifrequency, high-resolution radio recombination line observations of the massive star-forming region W49A. *Astrophys. J.* **482**, 307–333 (1995).
- Morris, M. & Serabyn, E. The Galactic Center Environment. *Annu. Rev. Astron. Astrophys.* **34**, 645–701 (1996).
- Westerlund, B. E. *The Magellanic Clouds* 202–220 (Cambridge Univ. Press, 1997).
- Serabyn, E. & Güsten, R. A molecular counterpart of the Galactic center arc. *Astron. Astrophys.* **184**, 133–143 (1987).
- Arp, H. & Sandage, A. Spectra of the two brightest objects in the amorphous galaxy NGC 1569: superluminous young star clusters—or stars in a nearby peculiar galaxy? *Astronom. J.* **90**, 1163–1171 (1985).

22. Wainscoat, R. J. & Cowie, L. L. A filter for deep near-infrared imaging. *Astron. J.* **103**, 332–337 (1992).
 23. Stetson, P. B. DAOPHOT: a computer program for crowded-field stellar photometry. *Publ. Astron. Soc. Pacif.* **99**, 191–222 (1987).

Acknowledgements. We thank B. Schaeffer and W. Harrison for their assistance at the telescope. Data presented here were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership between the California Institute of Technology, the University of California, and NASA. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. This work was performed in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Correspondence and requests for materials should be addressed to E.S. (e-mail: serabyn@tacos.caltech.edu).

Self-organized growth of nanostructure arrays on strain-relief patterns

Harald Brune, Marcella Giovannini, Karsten Bromann & Klaus Kern

Institut de Physique Expérimentale, EPF Lausanne, CH-1015 Lausanne, Switzerland

The physical and chemical properties of low-dimensional structures depend on their size and shape, and can be very different from those of bulk matter. If such structures have at least one dimension small enough that quantum-mechanical effects prevail, their behaviour can be particularly interesting. In this way, for example, magnetic nanostructures can be made from materials that are non-magnetic in bulk¹, catalytic activity can emerge from traditionally inert elements such as gold², and electronic behaviour useful for device technology can be developed^{3,4}. The controlled fabrication of ordered metal and semiconductor nanostructures at surfaces remains, however, a difficult challenge. Here we describe the fabrication of highly ordered, two-dimensional nanostructure arrays through nucleation of deposited metal atoms on substrates with periodic patterns defined by dislocations that form to relieve strain. The strain-relief patterns are created spontaneously when a monolayer or two of one material is deposited on a substrate with a different lattice constant. Dis-

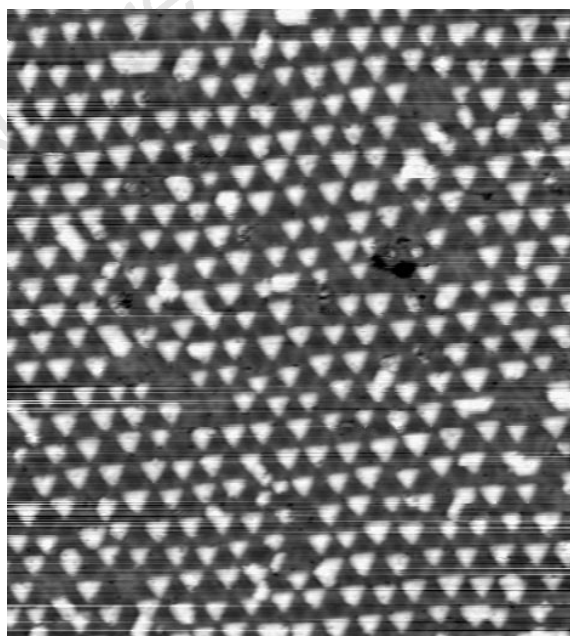


Figure 1 STM image of a periodic array of Fe islands nucleated on the dislocation network of a Cu bilayer on Pt(111) at 250 K.

locations often repel adsorbed atoms diffusing over the surface, and so they can serve as templates for the confined nucleation of nanostructures from adatoms. We use this technique to prepare ordered arrays of silver and iron nanostructures on metal substrates.

Arbitrary atomic-scale structures can be assembled with the tip of a scanning tunnelling microscope (STM), either through direct displacement of adsorbed atoms⁵, or through tip-assisted decomposition of chemical species⁶. The principal drawback of methods based on scanning probes is their serial character. Approaches where a large number of structures can be created in parallel are self-organized growth—either in the kinetic⁷ or in the thermodynamic regime^{8–10}—and the controlled deposition of size-selected clusters from the gas phase¹¹. Whereas the latter technique produces nanostructures of nearly uniform sizes on surfaces, the self-organized growth suffers from broad size distributions. In addition, both methods yield largely uncorrelated spatial distributions caused by the statistics of deposition and diffusion.

There have been several attempts to improve the spatial order in nanostructure growth. The vertical correlation of island nucleation in sequences of quantum dot and spacer layers yielded improved lateral order¹². Also, misfit dislocations have been used for nanoscale structuring. Preferred nucleation of Ni at dislocations of the Au(111) reconstruction resulted in ordered islands¹³; for this system, the mechanism was identified as site-specific exchange, a finding that strongly reduces the number of elements suitable for this type of ordering¹⁴. Also in semiconductors, island accumulation at dislocations has been reported¹⁵. However, there the bulk-like dislocations were not mobile enough to order into periodic patterns. Due to the attraction towards dislocations, islands were lined up in one dimension but were not periodic in two dimensions.

In heteroepitaxial systems that show strain relief through dislocations, the dislocations quite often arrange into highly ordered periodic patterns. This is due to mutual long-range repulsion and

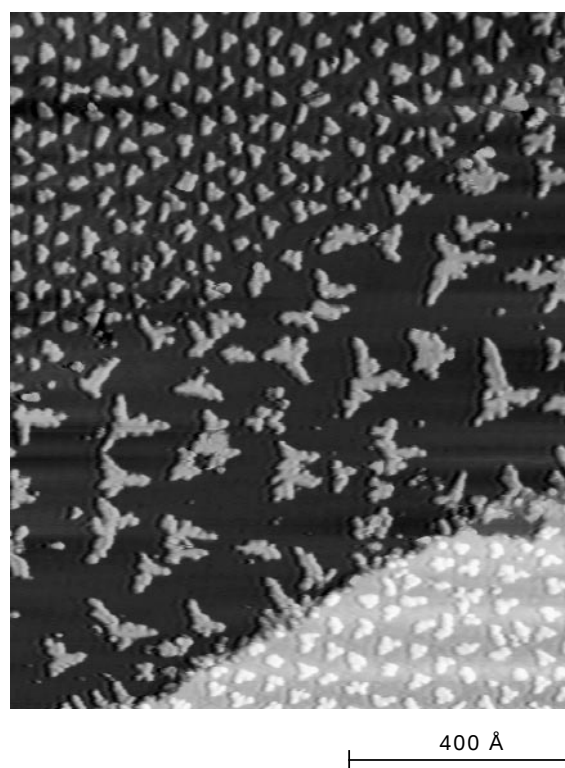


Figure 2 STM image of periodically ordered versus randomly nucleated islands on a heterogeneous substrate. The islands were grown by deposition of 0.1 ML Ag at 110 K onto a Pt(111) surface pre-covered by 1.5 ML Ag (1 ML is one adsorbed atom per substrate atom).

high mobility of the dislocations located at the surface. These surface dislocations are often strongly repulsive towards diffusing adatoms. Here we show that this can be used to transfer the dislocation periodicity through heterogeneous nucleation to regular superlattices of almost mono-dispersed islands. We demonstrate the feasibility of this self-organized growth method with Fe islands on the second monolayer (ML) of Cu on a Pt(111) substrate (Fig. 1), and with Ag nucleation on 2 ML of Ag on Pt(111) (Figs 2 and 3). The array of Fe quantum dots in Fig. 1 shows the unique potential of the approach. The areal density of the Fe dots exceeds the densities at present achieved by electron-beam lithography by two orders of magnitude.

Random versus ordered nucleation are contrasted in the STM image shown in Fig. 2. Quite regularly spaced islands coexist with a part of the surface covered by randomly distributed islands. While the latter show the broad distribution of island size (Fig. 3c) and distance typically observed for nucleation on isotropic substrates, the former appear to be placed on a periodic grid. This marked difference is caused by the structure of the underlying substrate. Before growing the Ag islands, the Pt(111) substrate was pre-

covered by 1.5 ML Ag such that the first and second monolayer coexist. These two layers differ in structure. The lattice constant of Ag is 4.3% larger than that of Pt. In the first monolayer, the Ag atoms are coherently strained by this amount and form a pseudomorphic layer¹⁶. As this layer is homogeneous (and isotropic), one observes random nucleation on top of it.

The second monolayer of Ag, on the other hand, forms a trigonal network of dislocations in which the compressive strain is partially relieved¹⁷. This network reveals a well established long-range order (Fig. 3a). The dislocations mark transitions from face-centred cubic (f.c.c.) to hexagonal close-packed (h.c.p.) stacking. Larger areas with the energetically favoured f.c.c. stacking are achieved by displacing one class of domain walls (Fig. 3a inset) relative to the crossing point of the two others. The resulting unit cell has trigonal symmetry and consists of a large quasi-hexagon (f.c.c.) and two triangles with opposite orientation (h.c.p. stacking).

Nucleation of Ag onto this network at low temperature leads to a high density of islands (Fig. 4), most of which are located away from dislocations. This implies that dislocations are repulsive barriers towards diffusing adatoms. At $T = 110$ K, the island density approaches a stationary value of exactly one island per network unit cell (Fig. 3b). We note that all islands nucleate within the distorted hexagons (Fig. 3d). This implies preferential binding to f.c.c.-areas in agreement with theory¹⁸. Owing to this difference in binding energy, atoms landing on h.c.p. triangles can diffuse into the f.c.c. hexagons but not vice versa. The repulsive nature of the dislocations and the attraction towards specific sites within the unit cell are the key properties transferring the periodicity of the dislocation network to a highly ordered two-dimensional island superlattice.

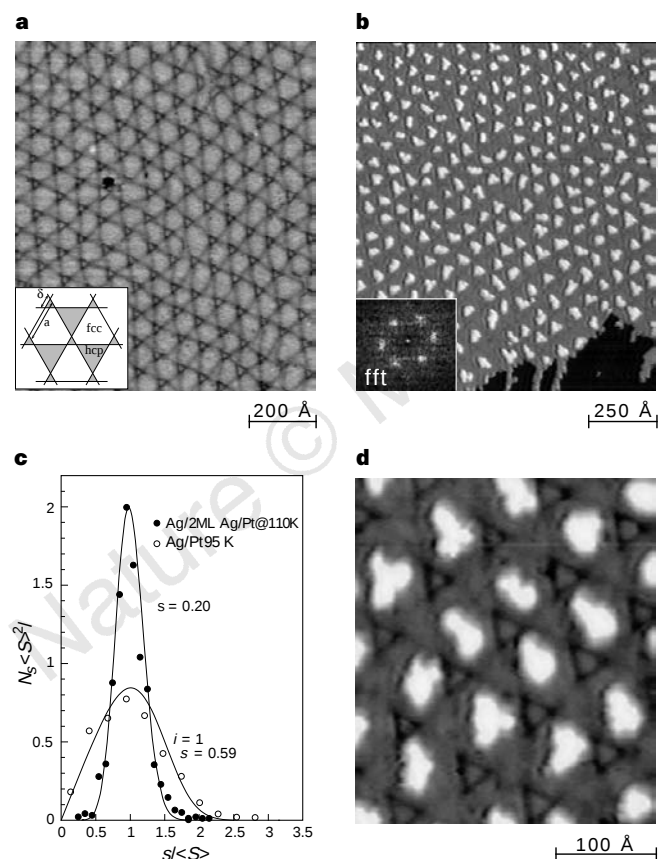


Figure 3 STM images showing the confined nucleation of adatom islands on a dislocation network. **a**, STM image of the ordered (25×25) dislocation network formed by the second Ag monolayer on Pt(111) on deposition at 400 K and subsequent annealing to 800 K. The inset shows a model of this trigonal strain-relief pattern. **b**, A superlattice of islands is formed on Ag deposition onto this network at 110 K (coverage $\theta = 0.10$ ML). Inset, the Fourier transform of the STM image shows the high degree of order and the hexagonal symmetry of the nanostructure array. **c**, Island size distributions for random and ordered nucleation (curves, theory; dots, experiment). The curve for ordered nucleation is a binomial fit. The curve labelled $i = 1$ shows the size distribution from scaling theory for random nucleation on an isotropic substrate. Size distributions were normalized according to scaling theory (s is the island size in atoms, $\langle S \rangle$ its mean value, and N_s the density of islands with size s per substrate atom). **d**, Detail of STM image **b**.

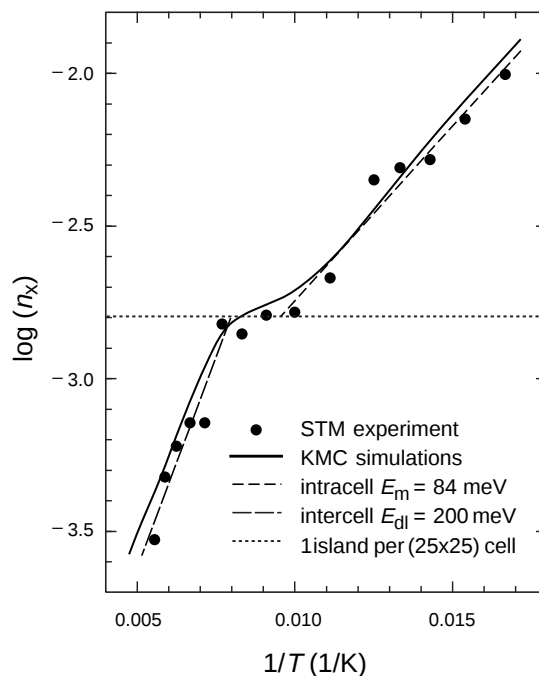


Figure 4 Arrhenius plot of measured (dots) and simulated (curve) island densities, n_x , for nucleation of 0.1 ML Ag on the dislocation network of Fig. 3a. The experimental island densities clearly show two slopes separated by a plateau with a density of one island per superstructure unit cell. The plateau is less pronounced in the Monte Carlo simulations than in the experiment. In the experiment, dislocations have a long-range repulsion towards adatoms, funneling them to the unit-cell centre where the islands are nucleating. We did not attempt to simulate the complicated energy surface required to account for these findings. The islands in the simulation are therefore randomly placed within the unit cells; those located at the border sometimes overgrow a dislocation which smears out the plateau.

The ordered nucleation is accompanied by an enhanced size uniformity (Fig. 3c). Atoms deposited onto a network unit cell stay confined to this cell by the surrounding dislocations. As all atoms arriving within one unit cell nucleate exactly one island there, the size distribution of these islands is binomial, $P_k = \binom{n}{k} p^k q^{n-k}$ (with $p = \Theta$, $q = 1 - \Theta$, k being the island size, n the size of the unit cell, and Θ the coverage). For our case of $p = 0.1$ and $n = 625$, the binomial distribution has a standard deviation of $\sigma = (qnp)^{1/2} = 0.12$ (normalized as in Fig. 3c). Owing to the convolution of tip and island shape there is a residual width explaining our larger experimental value of $\sigma = 0.20$. This value represents an upper bound to the real size distribution. We note that σ is often given for island diameters, where it is half the value of the island area distribution. Therefore our size distributions compare favourably to state-of-the-art results from self-organized quantum dots¹⁹.

Quantitative understanding of nucleation on the dislocation network is achieved through analysis of the Arrhenius behaviour of the saturation island density (Fig. 4). The system can be described very well by a transition between two rate-limiting processes, the intra-cell diffusion between adjacent lattice sites at low temperatures, and the inter-cell diffusion across dislocations at high temperatures. To analyse the slopes in Fig. 4 in terms of activation energies for these diffusion processes, we performed kinetic Monte Carlo simulations on a hexagonal lattice with a (25×25) superstructure. For simplicity we did not distinguish between f.c.c. and h.c.p. sites, so the unit cell was a simple rhombus. The free parameters are the energy barriers for diffusion within (E_m) and across (E_{dl}) the unit cells. The simulation clearly shows the transition between two slopes. It yields $E_m = 84 \pm 8$ meV for intra-cell diffusion and $E_{dl} = 200 \pm 30$ meV for diffusion across dislocations. The strong repulsion at dislocations explains the extended temperature regime of ordered nucleation. These results are, to our knowledge, the first quantitative measurements of diffusion on an inhomogeneous substrate.

For our approach of confined nucleation to work, the dislocations must arrange into periodic patterns, and they must be repulsive for adatom diffusion. The first condition is met by numerous heteroepitaxial systems. Besides the presented examples, well ordered trigonal dislocation networks are found in a number of epitaxial metal systems²⁰ and in semiconductors, for example, for surfactant-mediated growth of Ge on Si(111) (ref. 21). The requirement for dislocations to be repulsive is found to be met by many metal systems where there is no exchange²⁰. Also, moiré structures^{22–24} could be appropriate for the present approach of self-organization. Like dislocation patterns, they reveal periodic variations in adatom binding energy.

Our experiments describe a method to create almost monodispersed, equally spaced nanostructures through self-organization on substrates with a periodic strain-relief pattern. The inhomogeneous energy surface experienced by diffusing adatoms gives rise to heterogeneous nucleation and growth at specific sites within the superstructure, yielding well-ordered island superlattices. Our method may be useful for the growth of quantum arrays, that is, superlattices of nanostructures with size and period smaller than the Fermi wavelength. □

Received 23 December 1997; accepted 23 April 1998.

- Wildberger, K., Stepanyuk, V. S., Lang, P., Zeller, R. & Dederichs, P. H. Magnetic nanostructures: 4d clusters on Ag(001). *Phys. Rev. Lett.* **75**, 509–512 (1995).
- Haruta, M. Size- and support-dependency in the catalysis of gold. *Catal. Today* **36**, 153–166 (1997).
- Orlov, A. O., Amlani, I., Bernstein, G. H., Lent, C. S. & Snider, G. L. Realization of a functional cell for quantum-dot cellular automata. *Science* **277**, 928–930 (1997).
- Eberl, K. Quantum-dot lasers. *Phys. World* **10**(9), 45–50 (1997).
- Eigler, D. M. & Schweizer, E. K. Positioning single atoms with a scanning tunnelling microscope. *Nature* **344**, 524–526 (1990).
- Kent, A. D., Molnar, S., Gider, S. & Awschalom, D. D. Properties and measurement of scanning tunneling microscope fabricated ferromagnetic particle arrays. *J. Appl. Phys.* **76**, 6656–6660 (1994).
- Röder, H., Hahn, E., Brune, H., Bucher, J. P. & Kern, K. Building one- and two-dimensional nanostructures by diffusion-controlled aggregation at surfaces. *Nature* **366**, 141–143 (1993).
- Leonard, D., Krishnamurthy, M., Reaves, C. M., Denbaars, S. P. & Petroff, P. M. Direct formation of

- quantum-sized dots from uniform coherent islands of InGaAs on GaAs-surfaces. *Appl. Phys. Lett.* **63**, 3203–3205 (1993).
- Nötzel, R., Tommyo, J. & Tamamura, T. Self-organized growth of strained InGaAs quantum disks. *Nature* **369**, 131–133 (1994).
- Moison, J. M. et al. Self-organized growth of regular nanometer-scale InAs dots on GaAs. *Appl. Phys. Lett.* **64**, 196–198 (1994).
- Bromann, K. et al. Controlled deposition of size-selected Ag nanoclusters. *Science* **274**, 956–958 (1996).
- Tersoff, J., Teichert, C. & Lagally, M. G. Self-organization in growth of quantum dot superlattices. *Phys. Rev. Lett.* **76**, 1675–1678 (1996).
- Chambliss, D. D., Wilson, R. J. & Chiang, S. Nucleation of ordered Ni island arrays on Au(111) by surface-lattice dislocations. *Phys. Rev. Lett.* **66**, 1721–1724 (1991).
- Meyer, J. A., Baikie, J. D., Kopatzki, E. & Behm, R. J. Preferential island nucleation at the elbows of the Au(111) herringbone reconstruction through place exchange. *Surf. Sci.* **365**, L647–L651 (1996).
- Shiryaev, S. Y., Jensen, F., Hansen, J. L., Petersen, J. W. & Larsen, A. N. Nanoscale structuring by misfit dislocations in Si_{1-x}Ge_x/Si epitaxial systems. *Phys. Rev. Lett.* **78**, 503–506 (1997).
- Brune, H. et al. The effect of strain on surface diffusion and nucleation. *Phys. Rev. B* **52**, R14380–R14383 (1995).
- Brune, H., Röder, H., Boragno, C. & Kern, K. Strain relief at hexagonal close packed interfaces. *Phys. Rev. B* **49**, 2997–3000 (1994).
- Ratsch, C., Seitsonen, A. P. & Scheffler, M. Strain-dependence of surface diffusion: Ag on Ag(111) and Pt(111). *Phys. Rev. B* **55**, 6750–6753 (1997).
- Fafard, S., Leon, R., Leonard, D., Marz, J. L. & Petroff, P. M. Phonons and radiative recombination in self-assembled quantum dots. *Phys. Rev. B* **52**, 5752–5755 (1995).
- Brune, H. Kinetics of nucleation and aggregation in metal epitaxy. *Surf. Sci. Rep.* **31**, 121–126 (1998).
- Horn-von-Hoegen, M., Falou, A. A., Pietsch, H., Müller, B. H. & Henzler, M. Formation of the interfacial dislocation network in surfactant-mediated growth of Ge on Si(111) by SPALEED Part I. *Surf. Sci.* **298**, 29–42 (1993).
- Wiederholt, T., Brune, H., Wintterlin, J., Behm, R. J. & Ertl, G. Formation of two-dimensional sulfide phases on Al(111): an STM study. *Surf. Sci.* **324**, 91–105 (1994).
- Besenbacher, F., Nielsen, L. P. & Sprunger, P. T. in *Growth and Properties of Ultrathin Epitaxial Layers* (eds King, D. A. & Woodruff, D. P.) 207–257 (Elsevier, Amsterdam, 1997).
- Böhlinger, M., Jiang, Q., Berndt, R., Schneider, W. D. & Zegenhagen, J. Discommensurations, epitaxial growth and island formation in Cu/Ge(111). *Surf. Sci.* **367**, 245–260 (1996).

Acknowledgements. We thank J. Jacobsen, K. Jacobsen and J. Nørskov for providing us with the kinetic Monte Carlo computer code that we adopted for nucleation on a superlattice.

Correspondence and requests for materials should be addressed to K.K.

Doubling the critical temperature of La_{1.9}Sr_{0.1}CuO₄ using epitaxial strain

J.-P. Locquet*, J. Perret†, J. Fompeyrine‡, E. Mächler*, J. W. Seo*† & G. Van Tendeloo§

* IBM Research Division, Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland

† Institut de Physique, Université de Neuchâtel, CH-2000 Neuchâtel, Switzerland

‡ Institute of Inorganic Chemistry, University of Bern, CH-3012 Bern, Switzerland

§ EMAT, RUCA, University of Antwerp, B-2020 Antwerpen, Belgium

The discovery¹ of high-temperature superconductivity in copper oxides raised the possibility that superconductivity could be achieved at room temperature. But since 1993, when a critical temperature (T_c) of 133 K was observed in the HgBa₂Ca₂Cu₃O_{8+δ} (ref. 2), no further progress has been made in raising the critical temperature through material design. It has been shown, however, that the application of hydrostatic pressure can raise T_c —up to ~164 K in the case of HgBa₂Ca₂Cu₃O_{8+δ} (ref. 3). Here we show, by analysing the uniaxial strain and pressure derivatives of T_c , that compressive epitaxial strain in thin films of copper oxide superconductors could in principle generate much larger increases in the critical temperature than obtained by comparable hydrostatic pressures. We demonstrate the experimental feasibility of this approach for the compound La_{1.9}Sr_{0.1}CuO₄, where we obtain a critical temperature of 49 K in strained single-crystal thin films—roughly double the bulk value of 25 K. Furthermore, the resistive behaviour at low temperatures (but above T_c) of the strained samples changes markedly, going from insulating to metallic.

A high-pressure state in thin films can be induced by using epitaxial strain^{4,5}, but only moderate increases of T_c have been reported (+15%; refs 6, 7). Here we first re-examine published uniaxial pressure or strain derivatives of T_c , and predict that under compressive epitaxial strain a much larger increase in T_c must be

Table 1 Uniaxial pressure and strain derivatives with estimated possible increase of T_c

Compound	$\frac{\delta T_c}{\delta \epsilon_{ab}}$ (K)	$\frac{\delta T_c}{\delta \epsilon_c}$ (K)	ΔT_c^* (K)	$\frac{\delta T_c}{\delta P_{ab}}$ (K GPa ⁻¹)	$\frac{\delta T_c}{\delta P_c}$ (K GPa ⁻¹)	$\Delta T_c^\dagger$ (K)
La _{1.91} Sr _{0.09} CuO ₄ (ref. 8)	284	-851	10	3.2	-6.6	32
La _{1.9} Sr _{0.10} CuO ₄ (ref. 9)	375	-2,440	22	6.5	-13.8	65
YBa ₂ Cu ₃ O ₇ (ref. 10)				-0.05	-0.3	-0.5
YBa ₂ Cu ₃ O ₇ (ref. 11)				0.15	~0	1.5
YBa ₂ Cu ₃ O ₇ (ref. 5)				0.5	-	5
GdBa ₂ Cu ₃ O _{7-δ} (ref. 4)	-31	239	-1.2			
Bi ₂ Sr ₂ CaCu ₂ O ₈ (ref. 12)				1.8	-2.8	18
Bi ₂ Sr ₂ CaCu ₂ O ₈ (ref. 13)				1.3	-0.8	13
(Bi,Pb) ₂ Sr ₂ CaCu ₂ O ₈ (ref. 14)				10.0	-18.5	100
(Bi,Pb) ₂ Sr ₂ Ca ₂ Cu ₃ O ₁₀ (ref. 14)				6.2	-18.5	62

Values are shown for various compounds. For orthorhombic compounds, values along the *a* and *b* axes have been averaged.

* Estimated change in T_c for $\epsilon_{ab} = -\epsilon_c = 0.7\%$.

† Estimated change in T_c for $\Delta P = 5$ GPa.

possible. We then apply this idea to an underdoped La_{2-x}Sr_xCuO₄ (referred to here as '214', with *x* = 0.10) compound, which has a bulk T_c of 25 K.

The pressure (*P*) and strain ($\epsilon = [d_{\text{bulk}} - d_{\text{strained}}]/d_{\text{bulk}}$, with *d* a lattice parameter) dependence of T_c for a tetragonal unit cell contains an in-plane (*ab*) and an out-of-plane (*c*) term, and is written as follows:

$$T_c = T_c(0) + 2 \frac{\delta T_c}{\delta P_{ab}} P + \frac{\delta T_c}{\delta P_c} P$$

$$= T_c(0) + 2 \frac{\delta T_c}{\delta \epsilon_{ab}} \epsilon_{ab} + \frac{\delta T_c}{\delta \epsilon_c} \epsilon_c \quad (1)$$

Table 1 summarizes values of the uniaxial pressure and strain coefficients⁸⁻¹⁴. For most systems, the in-plane and out-of-plane T_c derivatives have opposite signs. Under hydrostatic pressure both terms nearly cancel, leading to a small increase of T_c (refs 15, 16). But under epitaxial growth conditions, ϵ_{ab} and ϵ_c normally have opposite signs (although this is not always the case¹⁷), and both terms add up to increase (compressive, $\epsilon_{ab} > 0$) or decrease (tensile, $\epsilon_{ab} < 0$) T_c .

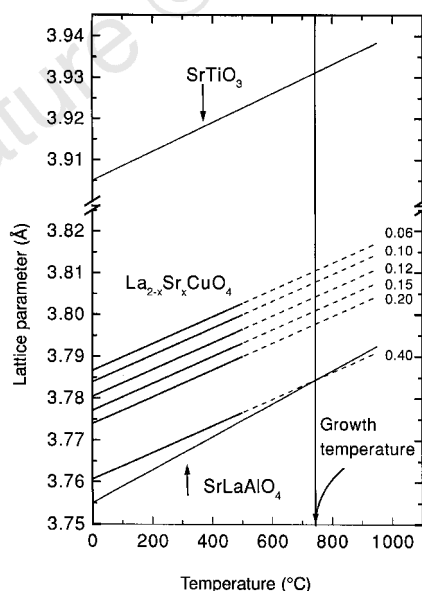


Figure 1 Extrapolated in-plane lattice parameters as a function of temperature. Values are shown for different Sr concentrations of the '214' compound (La_{2-x}Sr_xCuO₄, $\alpha = 8.5 \times 10^{-6} \text{ K}^{-1}$) as well as for the SLAO (SrLaAlO₄, $\alpha = 10.5 \times 10^{-6} \text{ K}^{-1}$) and STO (SrTiO₃, $\alpha = 9 \times 10^{-6} \text{ K}^{-1}$) substrates. Above 500 °C, α_{214} has not been reported. The reflection high-energy electron diffraction data give only relative measurements, but indicate that at 750 °C growth is compressive for *x* ≤ 0.12.

To increase T_c significantly, a large amount of compressive strain must be induced. This requires (1) the choice of a suitable system and substrate combination ('214' and SrLaAlO₄, referred to here as 'SLAO', with *a* = 3.754 Å); (2) a defect-free growth process (block-by-block molecular beam epitaxy¹⁸), and (3) an optimal thickness (10–15 nm). This thickness is a compromise between avoiding the nucleation of misfit dislocations and minimizing the T_c reduction observed for ultrathin films^{19,20}. Indeed, dislocations appear when a critical thickness h_c has been exceeded²¹; h_c decreases with increasing mismatch, suggesting that a large strain requires very thin films. Such thin (15-nm) '214' films were grown simultaneously on SLAO and SrTiO₃ (referred to here as 'STO' with *a* = 3.905 Å, tensile strain) by block-by-block molecular beam epitaxy^{18,22} at 750 °C, where a sequence of two monolayers of (La,Sr)–O followed by one monolayer of Cu–O is repeated. High-quality thin films can thus be prepared, and strain-relaxation processes related to the incorporation of defects can be minimized.

Precise estimates of ϵ demand values for the thermal expansion coefficients α of '214' and SLAO. Unfortunately, α_{214} is not known above 500 °C. In Fig. 1, $\alpha_{214} = 8.5 \times 10^{-6} \text{ K}^{-1}$ (ref. 23) and $\alpha_{\text{SLAO}} = 10.5 \times 10^{-6} \text{ K}^{-1}$ were used, and compressive strain is predicted at the growth temperature for all '214' compounds with *x* ≤ 0.4. However, measurements of ϵ_{ab} during growth using reflection high-energy electron diffraction²² revealed compressive strain

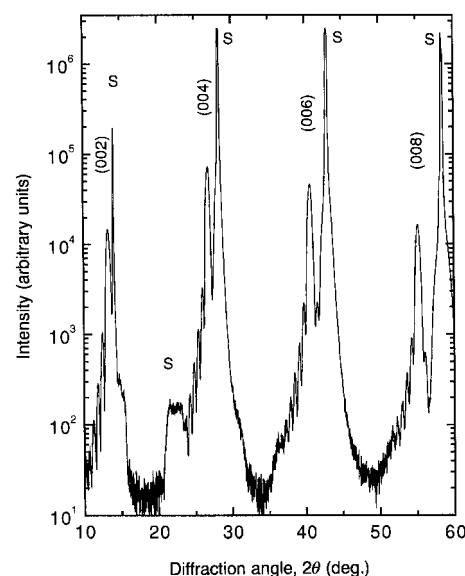


Figure 2 Measured X-ray diffraction pattern of the 12-unit-cell-thick '214' film on SLAO. Besides the substrate peaks indicated by S, only the '214' (00*l*) peaks with their finite-size oscillations are observed. The presence of these finite-size oscillations suggests a film roughness of ±(1–2) unit cells.

only up to $x = 0.12$. This discrepancy might be related to a smaller α_{214} at high temperatures or to the presence of an interface layer²⁴. In any case, a large compressive strain requires $x < 0.12$.

The θ - 2θ X-ray diffraction pattern of the film grown on SLAO reveals (00 l) peaks and finite-size oscillations (Fig. 2). From this and measurements taken around the (107), (109) and (1011) reflections, the lattice parameters $c = 13.31 \pm 0.01$ Å and $a = 3.76 \pm 0.02$ Å were determined (bulk compound²⁵: 13.212 and 3.784 Å, respectively) leading to $\epsilon_{ab} = 0.63\%$, $\epsilon_c = -0.76\%$. For the film grown on STO, similar measurements yield $\epsilon_{ab} = -0.54\%$, $\epsilon_c = 0.35\%$. Preliminary X-ray refinements of these data using a modified Suprex²⁶ program yield a change of the distance between two CuO₂ planes from 6.61 Å (bulk) to 6.65 Å (SLAO) and 6.59 Å (STO). The cross-section, high-resolution transmission electron microscopy image (Fig. 3) of the film grown on SLAO confirms the low defect density in the sample. The film-substrate interface, indicated by arrows, has steps that are one-half of a unit cell high. Despite these steps, perfect continuity of the first '214' layer was obtained, and no defects were observed. On the basis of image simulations (see Fig. 3 inset), the stacking sequence of the interface can be described as LaO-LaO-AlO₂-LaO-CuO₂-LaO-LaO; that is, there is an intermediate La₂CuAlO₆ layer at the interface. In Fig. 3, the white dots in the film and substrate correspond to Cu and Al columns, respectively. By taking the Al-Al distance as a standard, the distances between two CuO₂ planes are obtained directly from lattice fringe spacing refinements, namely 6.63 Å (SLAO) and 6.57 Å (STO). A few misfit dislocations only were found in plan-view samples with an average distance of 200 nm. These have an average length of 150 nm and Burgers vectors of $\mathbf{b} = \langle 100 \rangle$ and $\langle 110 \rangle$. Their presence suggests that the film has nevertheless partially relaxed; thinner films might allow us to increase the strain further.

The values of ϵ determined above, together with the $\delta T_c/\delta \epsilon$ coefficients of Table 1 and the phenomenological equation (1), suggest a substantial change of T_c . Measurements of the resistivity versus temperature $\rho(T)$ (Fig. 4) indeed reveal that the film grown under compressive strain has a T_c of 49.1 K, while the film grown under tensile strain has a T_c of 10 K. Compared with the T_c of the bulk compound²⁷ (25 K, see dotted line in Fig. 4), these values represent a spectacular increase and decrease, respectively. Such large changes have not been achieved for bulk '214', not even under the highest hydrostatic or uniaxial pressure used up to now.

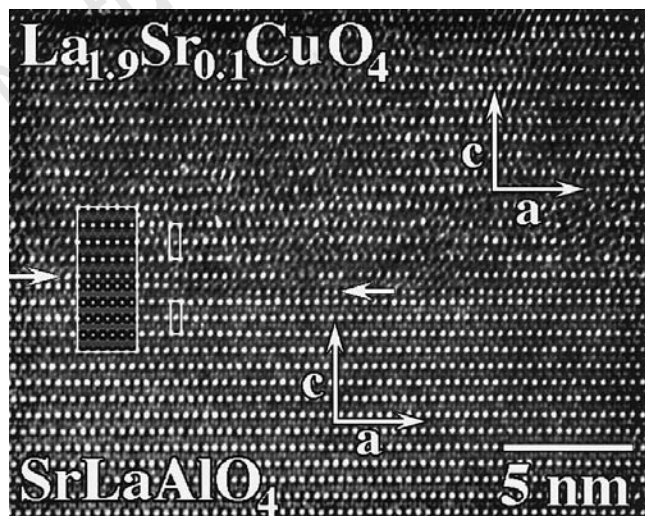


Figure 3 Cross-section, high-resolution electron-microscopy image of the '214' film on SLAO. On the basis of image simulations, the white dots in the film and substrate can be correlated with Cu and Al columns, respectively. The unit cells are indicated by the small white rectangles. The inset shows the simulated image of the interface structure with the stacking sequence LaO-LaO-AlO₂-LaO-CuO₂-LaO-LaO.

At low temperatures ($T < 100$ K), another significant but unexpected change occurs. Compared to the resistive behaviour of the bulk compound (as shown by the $\rho(T)$ curve), the film grown under tensile strain shows a large upturn in this curve as temperature is reduced, before the onset of superconductivity. Such a behaviour has been observed for heavily underdoped '214' compounds ($x \approx 0.06$; ref. 27). On the other hand, the film grown under compressive strain shows a metallic and linear $\rho(T)$, which is usually observed for optimally doped compounds ($x \approx 0.16$; ref. 27). However, we emphasize that in both cases the overall $\rho(T)$ behaviour remains very different from that of typical underdoped or optimally doped samples. For instances, the ratio $\rho(300 \text{ K})/\rho(50 \text{ K})$ for the film on SLAO is ~ 8 , compared with ~ 4 for the best optimally doped bulk samples²⁷. Moreover, at relatively high temperatures (270 K), a change of slope $\delta\rho/\delta T$ occurs for the film on SLAO, towards a value typically observed above the 'pseudogap' temperature T^* (see dashed line in Fig. 4). For the bulk compound, on the other hand, such a change of slope occurs at $T^* \approx 600$ K (ref. 28); that is, under compressive strain T^* is reduced by a factor of two. Hence, these data suggest that in addition to causing a substantial increase of T_c , strain also significantly modifies the normal-state properties.

These changes could partially be explained by variations of the carrier density n . Indeed as n increases, an insulator-to-metal transition is observed²⁷. A correlation between an increasing T_c and decreasing T^* as n becomes larger is commonly reported for other copper oxides²⁹. However, a rough estimate using the values of $\rho(300 \text{ K})$ indicates that these variations are not larger than $\pm 10\%$. This certainly cannot explain the observed T_c value of 49 K, which is significantly higher than the maximum observed in the bulk '214' system (37 K for $x = 0.16$).

These results give rise to two fundamental questions. First, can the T_c of other compounds be increased? Table 1 shows that under compressive strain ($P = 5 \text{ GPa}$ or $\epsilon_{ab} = -\epsilon_c = 0.7\%$), significant increases can be expected for some compounds. Unfortunately the uniaxial T_c derivatives for the Tl and Hg compounds are not yet available. Second, how do these observations change our understanding of superconductivity in copper oxides? If the changes in n

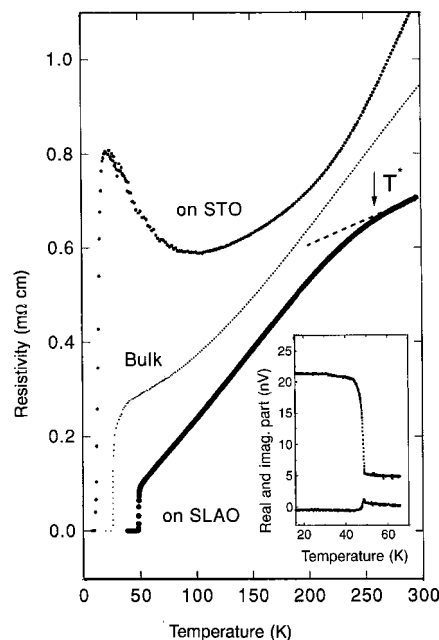


Figure 4 Resistivity versus temperature for the two films grown simultaneously and for the bulk $x = 0.10$ sample²⁶. Data were obtained using a four-point method. Inset, Real and imaginary parts of the a.c. susceptibility of the film on SLAO, which reveals a sharp, single-phase transition $\Delta T_c = 0.7$ K; the transition width of the film on STO is $\Delta T_c = 4$ K.

amount to $\pm 10\%$, this would result in a total T_c variation of 7 K in the bulk phase diagram. In strained films a total variation of ~ 40 K is observed, together with a drastic change in T^* . Hence, the lattice deformations associated with the strain fundamentally modify the energy scales, leading to the formation and condensation of the superconducting pairs. As the pair formation strongly influences the spin and charge excitation spectrum of the normal state, $\rho(T)$ and the metal–insulator transition are also considerably modified. Further work to identify and quantify the subtle lattice deformations responsible for these changes needs to be done. In any case, the data reported here emphasize that T_c increases as the distance between consecutive CuO_2 planes increases, in contradiction of a recent theoretical prediction³⁰. Our data also show that adjustment of the epitaxial strain offers a unique way to increase T_c and to reveal the fundamental role played by the interlayer coupling. $\square$

Received 13 March; accepted 27 May 1998.

- Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba–La–Cu–O system. *Z. Phys.* **B 64**, 189–193 (1986).
- Schilling, A., Cantoni, M., Guo, J. D. & Ott, H. R. Superconductivity above 130 K in the Hg–Ba–Ca–Cu–O system. *Nature* **362**, 56–58 (1993).
- Gao, L. *et al.* Superconductivity up to 164 K in $\text{HgBa}_2\text{Ca}_{m-1}\text{Cu}_m\text{O}_{2m+2+\delta}$ ($m = 1, 2$ and 3) under quasihydrostatic pressures. *Phys. Rev. B* **50**, 4260–4263 (1994).
- Bud'ko, S. L., Guimpel, J., Nakamura, O., Maple, M. B. & Schuller, I. K. Uniaxial pressure dependence of the superconducting critical temperature in $\text{RBa}_2\text{Cu}_3\text{O}_{7-\delta}$ high T_c oxides. *Phys. Rev. B* **46**, 1257–1260 (1992).
- Belenky, G. L. *et al.* Effect of stress along the *ab* plane on the J_c and T_c of $\text{YBa}_2\text{Cu}_3\text{O}_7$ thin films. *Phys. Rev. B* **44**, 10117–10120 (1991).
- Sato, H. & Naito, M. Increase in the superconducting transition temperature by anisotropic strain effect in $(001) \text{La}_{1.8}\text{Sr}_{0.2}\text{CuO}_4$ thin films on LaSrAlO_4 substrates. *Physica C* **274**, 221–226 (1997).
- Locquet, J.-P. & Williams, E. J. Epitaxially induced defects in Sr- and O-doped La_2CuO_4 thin films grown by MBE: implications for transport properties. *Acta Polonica A* **92**, 69–84 (1997).
- Nohara, M. *et al.* Unconventional lattice stiffening in superconducting $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ single crystals. *Phys. Rev. B* **52**, 570–580 (1995).
- Gugenberger, F. *et al.* Uniaxial pressure dependence of T_c from high resolution dilatometry of untwinned $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ single crystals. *Phys. Rev. B* **49**, 13137–13142 (1994).
- Welp, U. *et al.* Effect of uniaxial stress on the superconducting transition in $\text{YBa}_2\text{Cu}_3\text{O}_7$. *Phys. Rev. Lett.* **69**, 2130–2133 (1992).
- Meingast, C. *et al.* Large *a*–*b* anisotropy of the expansivity anomaly at T_c in untwinned $\text{YBa}_2\text{Cu}_{7-\delta}$. *Phys. Rev. Lett.* **67**, 1634–1637 (1991).
- Meingast, C., Junod, A. & Walker, E. Superconducting fluctuations and uniaxial pressure dependence of T_c of a $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ single crystal from high resolution thermal expansion. *Physica C* **272**, 106–114 (1996).
- Fukamachi, K. *et al.* in *Advances in Superconductivity* Vol. VII (eds Yamafuji, K. & Morishita, T.) 225–228 (Springer, Tokyo, 1995).
- Chen, X., Tessema, G. X. & Skove, M. J. Effects of elastic stress on the resistivity and T_c of $(\text{Bi,Pb})_2\text{Sr}_2\text{Ca}_{n-1}\text{Cu}_n\text{O}_x$. *Physica C* **181**, 340–344 (1991).
- Tanahashi, N. *et al.* The strontium content dependence of pressure effect in $(\text{La}_{1-x}\text{Sr}_x)_2\text{CuO}_4$. *Jpn J. Appl. Phys.* **28**, L762–L765 (1989).
- Yamada, N. & Ido, M. Pressure effects on superconductivity and structural phase transitions in $\text{La}_{2-x}\text{M}_x\text{CuO}_4$ ($M = \text{Ba}, \text{Sr}$). *Physica C* **203**, 240–246 (1992).
- Fartash, A., Grimsditch, M., Fullerton, E. E. & Schuller, I. K. Breakdown of Poisson's effect in Nb/Cu superlattices. *Phys. Rev. B* **47**, 12813–12819 (1993).
- Locquet, J.-P. & Mächler, E. Block-by-block deposition of complex oxide films. *Mater. Res. Soc. Bull.* **19**, 39–43 (1994).
- Triscone, J.-M. *et al.* $\text{YBa}_2\text{Cu}_3\text{O}_7/\text{PrBa}_2\text{Cu}_3\text{O}_7$ superlattices: Properties of ultrathin superconducting layers separated by insulating layers. *Phys. Rev. Lett.* **64**, 804–807 (1990).
- Terashima, T. *et al.* Superconductivity of one-unit-cell thick $\text{YBa}_2\text{Cu}_3\text{O}_7$ thin film. *Phys. Rev. Lett.* **67**, 1362–1365 (1991).
- Matthews, J. W. & Blakeslee, A. E. Defects in epitaxial multilayers. *J. Cryst. Growth* **27**, 118–125 (1994).
- Locquet, J.-P. *et al.* Variation of the in-plane penetration depth λ_{ab} as a function of doping in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_{4+\delta}$ thin films on SrTiO_3 : implications for the overdoped state. *Phys. Rev. B* **54**, 7481–7488 (1996).
- Tresse, F. *Cristallochimie de Quelques Oxydes des Éléments de Transition: Influence de la présence d'oxygène et de Substitutions Cationiques sur la Supraconduction* Thesis, Univ. Bordeaux (1990).
- Hojczyk, R., Jia, C.-L., Poppe, U. & Urban, K. New insulating materials for high T_c superconductor device applications. *Physica C* **282**–**287**, 731–732 (1997).
- Radaelli, P. G. *et al.* Structural and superconducting properties of $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ as a function of Sr content. *Phys. Rev. B* **49**, 4163–4175 (1994).
- Fullerton, E. E., Schuller, I. K., Vanderstraeten, H. & Bruynseraede, Y. Structural refinements from x-ray diffraction. *Phys. Rev. B* **45**, 9292–9310 (1992).
- Kimura, T. *et al.* In-plane and out-of-plane magnetoresistance in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ single crystals. *Phys. Rev. B* **53**, 8733–8742 (1996).
- Nakano, T. *et al.* Magnetic properties and electronic conduction of superconducting $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. B* **49**, 16000–16008 (1994).
- Williams, G. V. M., Tallon, J. L., Haines, E. M., Michalak, R. & Dupree, R. NMR evidence for a d-wave normal state pseudogap. *Phys. Rev. Lett.* **78**, 721–724 (1997).
- Anderson, P. W. *The Theory of Superconductivity in the High T_c Cuprate Superconductors* (Princeton Univ. Press, 1997); C-axis electrodynamic evidence for the interlayer theory of high temperature superconductivity. *Science* **279**, 1196–1198 (1998).

Acknowledgements. We thank Y. Jaccard, F. Arrouty and E. J. Williams for contributions in the initial phase of this project; P. Martinoli, O. Fischer, J.-M. Triscone, J. G. Bednorz and C. Rossel for discussions; and C. Voisard and T. Maeder for the thermal expansion measurements. This work was supported by the Swiss National Science Foundation, the Swiss Priority Project “Minast”, and the Belgian UIAP 4/10.

Correspondence and requests for materials should be addressed to J.-P.L. (e-mail: loc@zurich.ibm.com).

Nanocomposite polymer electrolytes for lithium batteries

F. Croce, G. B. Appetecchi, L. Persi & B. Scrosati

Dipartimento di Chimica, Sezione di Elettrochimica, Università “La Sapienza”, 00185 Rome, Italy

Ionically conducting polymer membranes (polymer electrolytes) might enhance lithium-battery technology by replacing the liquid electrolyte currently in use and thereby enabling the fabrication of flexible, compact, laminated solid-state structures free from leaks and available in varied geometries¹. Polymer electrolytes explored for these purposes are commonly complexes of a lithium salt (LiX) with a high-molecular-weight polymer such as polyethylene oxide (PEO). But PEO tends to crystallize below 60 °C, whereas fast ion transport is a characteristic of the amorphous phase. So the conductivity of PEO–LiX electrolytes reaches practically useful values (of about $10^{-4} \text{ S cm}^{-1}$) only at temperatures of 60–80 °C. The most common approach for lowering the operational temperature has been to add liquid plasticizers, but this promotes deterioration of the electrolyte's mechanical properties and increases its reactivity towards the lithium metal anode. Here we show that nanometre-sized ceramic powders can perform as solid plasticizers for PEO, kinetically inhibiting crystallization on annealing from the amorphous state above 60 °C. We demonstrate conductivities of around $10^{-4} \text{ S cm}^{-1}$ at 50 °C and $10^{-5} \text{ S cm}^{-1}$ at 30 °C in a PEO– LiClO_4 mixture containing powders of TiO_2 and Al_2O_3 with particle sizes of 5.8–13 nm. Further optimization might lead to practical solid-state polymer electrolytes for lithium batteries.

The basic structure of PEO–LiX polymer electrolytes involves PEO chains coiled around the Li^+ cations, thereby separating them from the X^- counteranions^{2,3}. Polymer electrolytes therefore require local relaxation and segmental motion of the solvent (PEO) chains to allow ion (Li^+) transport, and this condition can only be obtained when the polymer is in an amorphous state, that is, above 60 °C.

Large research efforts have been devoted to lowering to the ambient region the temperature of operation of the PEO–LiX polymer electrolytes. The most common approach has been that of adding liquid plasticizers, such as low-molecular-weight polyethyleneglycols or aprotic organic solvents, to the PEO–LiX matrix. But because of the problems mentioned above, the gain in conductivity is adversely accompanied by a loss of the solid-state configuration and by a loss of the compatibility with the lithium electrode; this represents the loss of the most important intrinsic features of the polymer electrolyte. So liquid plasticizer-added PEO–LiX electrolytes cannot be used in current lithium-metal batteries because their limited processability and high reactivity result in serious problems in terms of battery cyclability and safety.

It is therefore thought that only plasticizer-free electrolytes can assure an efficient cyclability of the lithium metal electrode. On the other hand, these electrolytes conduct only at temperatures above 60 °C. What is needed is some way to enhance the low-temperature ionic conductivity of PEO–LiX electrolytes without introducing a liquid phase. Our approach involves adding ceramic powders of nanoscale particle size, to develop solid-state PEO–LiX composite polymer electrolytes having, in the range 30–80 °C, both excellent mechanical stability (promoted by the network of the fillers into the polymer bulk) and high ionic conductivity (promoted by the high surface area of the dispersed fillers).

The general idea of adding ceramic powders to PEO–LiX electrolytes is not new. This procedure has been successfully employed

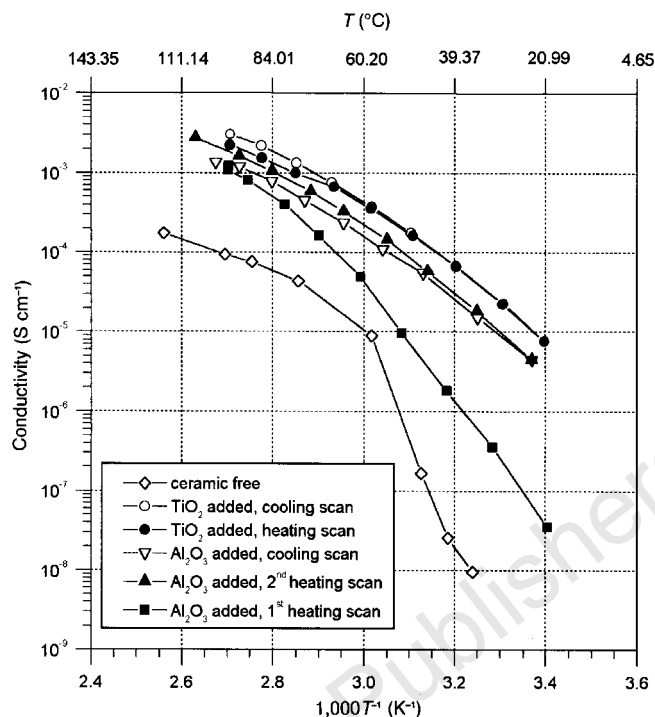


Figure 1 Arrhenius plots of the conductivity of ceramic-free PEO-LiClO₄ and of nanocomposite PEO-LiClO₄/10 wt% TiO₂ and PEO-LiClO₄/10 wt% Al₂O₃ polymer electrolytes (PEO : LiClO₄ = 8 : 1 in all cases). Data obtained by impedance spectroscopy measurements.

for improving the mechanical⁴ and the interfacial^{5–7} properties of the PEO-LiX electrolytes. However, to our knowledge, it has never been used for promoting low-temperature conductivity. The innovation here proposed is based on the selection of ceramics with nanoscale particle dimensions and suitable surface characteristics. Observations of conductivity enhancement following the addition of ceramic powders has been reported in the past by us^{8,9,19} and by others¹⁰, but the ceramic filler's effective role in promoting ion transport was never clearly identified. Raman spectroscopy studies have suggested that the structural properties of solid composite PEO-LiX electrolytes containing dispersed, low-particle-size ceramic fillers (for example, γ -LiAlO₂ fillers) might be comparable with those of liquid electrolytes formed by low-molecular-weight polyethylene glycol-lithium salt solutions¹¹. More recently, Wiczorek *et al.*^{12,13} have reported that a reduction below 4 μ m of the particle size of added ceramics (for example, Al₂O₃ ceramics) in composite PEO-NaI electrolytes was accompanied by an increase of ionic conductivity. These authors also suggested an active role of the surface groups of the ceramic particles in promoting local structural modifications. We also note a report¹⁴ of conductivity enhancement in PEO-LiX composite electrolytes to which ceramics (for example, TiO₂) have been added: however, this experimental result was not explained.

A careful examination of these early works leads to the conclusion that the dispersed ceramics influence the recrystallization kinetics of the PEO polymer chains, thereby ultimately promoting localized amorphous regions and thus enhancement of the Li⁺ ion transport. One could consider tailoring these structural modifications by selecting inorganic fillers having particular chemical properties and by maximizing their surface area. Following this concept, we describe here the development and characterization of composite electrolytes formed by dispersing into a PEO-LiClO₄ matrix ceramic powders (for example, titania or alumina) with particle size reduced to the nanoscale region. It is expected that, once these composite electrolytes are annealed at temperatures higher than the PEO crystalline-to-amorphous transition (that is, above 60 °C), the ceramic additive, due to the combined action of its large surface area and its Lewis-acid character, will prevent PEO chain reorganization: the expected result is the freezing at ambient temperature of a high degree of disorder, which is likely to be accompanied by a consistent enhancement of the ionic conductivity. The goal is to obtain a new class of solid polymer electrolytes having unique properties such as: (1) low-temperature conductivity (promoted by the high surface area of the dispersed filler), (2) improved mechanical stability (induced by the network of the dispersed filler) and (3) good compatibility with the lithium metal electrode (assured by the

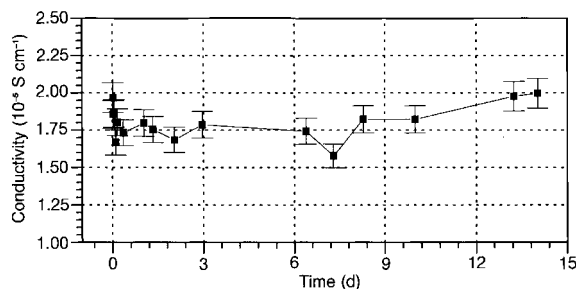


Figure 2 Change with time of the conductivity of the nanocomposite PEO-LiClO₄/10 wt% TiO₂ polymer electrolyte at 31 °C (PEO : LiClO₄ = 8 : 1).

absence of any liquids and by the interfacial stabilizing action of the dispersed filler).

The ceramic fillers (TiO_2 , 13 nm particle size (Degussa and CISE Laboratories, Milan) and Al_2O_3 , 5.8 nm particle size (Aldrich)) were dried at 250 °C for 24 hours before use. The preparation of the composite electrolyte samples involved first the dispersion of the selected ceramic powder and of the lithium salt (LiClO_4) in acetonitrile, followed by the addition of the polymer component (PEO) and by a thorough mixing of the resulting slurry. The LiClO_4 /PEO mole ratio was fixed at 1:8, and the amount of added ceramic was fixed at 10 wt% of the total PEO + LiClO_4 weight. The slurry was then cast between two glass plates separated by spacers. These procedures yielded homogenous and mechanically stable membranes of average thickness 100 μm .

Figure 1 compares the Arrhenius plot of a ceramic-free PEO– LiClO_4 polymer electrolyte with those of the nanocomposite PEO– LiClO_4 .10 wt% TiO_2 and PEO– LiClO_4 .10 wt% Al_2O_3 electrolytes. The heating scan of the ceramic-free electrolyte shows a break around 60 °C, reflecting the well-known transition from the crystalline state of PEO to the amorphous state, which is accompanied by an increase in ionic conductivity. The trend of the curve is reproduced in the following cooling scan; this confirms that, when brought back to low temperature, the common PEO– LiClO_4 electrolyte tends to recrystallize and, consequently, the conductivity decays. The as-prepared nanocomposite electrolytes have a room-temperature conductivity and a first heating scan similar to that of the plain electrolyte. However, the behaviour of the following cooling scan is quite different; no break occurs around 60 °C and the conductivity remains consistently higher (between 10^{-3} and $10^{-5} \text{ S cm}^{-1}$ versus the range 10^{-4} to $10^{-8} \text{ S cm}^{-1}$ for the ceramic-free electrolytes in the temperature range 80–30 °C. The conductivity trend of the composite electrolytes is reproduced in the following heating and cooling scans. Furthermore, the value at ambient temperature stays stable for several days, as shown by Fig. 2 where we report the conductivity of the PEO– LiClO_4 .10 wt% TiO_2 nanocomposite polymer electrolyte as a function of time at 31 °C. Eventually, recrystallization phenomena may occur, but with kinetics which appear to be critically dependent on the annealing conditions.

These results show convincingly the conductivity enhancement in the nanocomposite polymer electrolytes. Further confirmation has been obtained by a series of complementary tests. Infrared spectra have shown that the trace content of the residual casting solvent is similar in both the nanocomposite electrolytes and the ceramic-free electrolytes. This rules out the possibility that the conductivity enhancement in the former, rather than being due to a faster ion transport, could have been related to a dilution and/or plasticizing effect promoted by an excess of liquid adsorbed by the ceramics during casting and released in the electrolyte bulk during the initial heating scan. Differential scanning calorimetry carried out on as-prepared nanocomposite electrolyte samples have shown in the first heating scan a peak due to the crystalline-to-amorphous transition at 60–70 °C. However, no recrystallization peaks are observed in the following cooling scan (from 110 °C to room temperature) or in the subsequent heating scans of the same samples, even if kept at room temperature for several days. This confirms that the nanocomposite electrolytes retain their amorphous state on storage at room temperature. We note that stress-strain measurements showed a large enhancement of the Young's modulus and of the yield-point stress when passing from ceramic-free to nanocomposite polymer electrolytes, thereby demonstrating that the conductivity state of the latter is not due to polymer degradation but rather is accompanied by a substantial increase in the electrolyte's mechanical properties.

We have also determined the value of the lithium-ion transference number. In conventional, ceramic-free PEO–LiX electrolytes, the fraction of current transported by the Li^+ cations is lower than that transported by the X^- anions, being usually in the range 0.2–0.3 (ref.

15). We have determined the Li^+ ion transference number in nanocomposite PEO– LiClO_4 .10 wt% TiO_2 electrolytes using two independent methods; the method proposed by Vincent and co-workers¹⁶ and the refined method proposed by Abraham and co-workers¹⁷. The results give an average Li^+ transference number of the order of 0.6 in the 45–90 °C temperature range. To our knowledge, such a high value has not been previously reported for PEO–LiX electrolytes. This value, however, is consistent with a transport model based on the electrolyte's structural modifications induced via Lewis acid–base interactions between the ceramic surface states and both the X^- anions and the PEO segments¹³. These interactions originate a series of processes which work together to increase the fraction of free ions and to enhance their mobility. In fact, it is reasonable to assume that the acid sites on the surface of the ceramic particles may compete with the acid lithium cations to form complexes with the basic oxygens on the PEO chains. Thus the ceramics may act as cross-linking centres for PEO segments, thereby lowering the reorganization tendency of the polymer chain and promoting preferred Li^+ transport routes at the boundaries of the ceramic particles. Under these conditions, a consistent enhancement of the cation transference number is expected.

The high Li^+ transference number and the high and stable conductivity at moderate temperature are, to our knowledge, the first reported convincing experimental evidence to support previous hypotheses on the role of the ceramic filler in promoting enhanced Li^+ transport via a surface mechanism^{8–10,13,19}. We believe that the results obtained here could be optimized by a proper choice of the type and morphology of the ceramic filler. This might lead to the development of polymer electrolytes having a true solid-state configuration (and thus good mechanical properties) combined with high conductivity and, most importantly, with high lithium ion transference number and high interfacial stability with the lithium metal electrode. Electrolytes having these properties have been long searched for, as they are ideal candidates for use in rechargeable lithium batteries¹⁸. □

Received 4 February; accepted 1 June 1998.

- Scrosati, B. Lithium batteries. *Chim. Ind. (Milan)* **79**, 463–471 (1997).
- Lightfoot, P., Metha, M. A. & Bruce, P. G. Crystal structure of the polymer electrolyte poly(ethylene oxide)₃:LiClF₃. *Science* **262**, 883–885 (1993).
- Vincent, C. A. & Scrosati, B. C. H. *Modern Batteries. An Introduction to Electrochemical Power Sources* 2nd edn (Arnold, London, 1997).
- Weston, J. E. & Steele, B. C. H. Effects of inert fillers on the mechanical and electrochemical properties on lithium salt-poly(ethylene oxide) polymer electrolytes. *Solid State Ionics* **7**, 75–79 (1982).
- Croce, F. & Scrosati, B. Interfacial phenomena in polymer electrolyte cells: lithium passivation and cyclability. *J. Power Sources* **43**, 9–19 (1993).
- Borghini, M. C., Mastragostino, M., Passerini, S. & Scrosati, B. Electrochemical properties of poly(ethylene oxide)-Li[(CF₃SO₂)₂N]—gamma-LiAlO₂ composite polymer electrolytes. *J. Electrochem. Soc.* **142**, 2118–2121 (1995).
- Capuano, F., Croce, F. & Scrosati, B. Composite polymer electrolytes. US Patent No. 5,576,115 (1996).
- Croce, F., Bonino, F., Panero, S. & Scrosati, B. Properties of mixed polymer and crystalline ionic conductors. *Phil. Mag. B* **59**, 161–168 (1989).
- Capuano, F., Croce, F. & Scrosati, B. Composite polymer electrolytes. *J. Electrochem. Soc.* **52**, 1918–1922 (1991).
- Kumer, B. & Scanlon, Y. G. Polymer ceramic composite electrolytes. *J. Power Sources* **52**, 261–268 (1994).
- Croce, F., Scrosati, B. & Mariotto, G. Electrochemical and spectroscopic study of the transport properties of composite polymer electrolytes. *Chem. Mater.* **4**, 1134–1136 (1992).
- Wieczorek, W., Florjanczyk, Z. & Stevens, J. R. Composite polyether based electrolytes. *Electrochim. Acta* **40**, 2251–2258 (1995).
- Przyłuski, J., Siekierski, M. & Wieczorek, W. Effective medium theory in studies of conductivity of composite polymeric electrolytes. *Electrochim. Acta* **40**, 2101–2108 (1995).
- Kumer, B. paper presented at 5th Workshop for Battery Exploration Development, US Navy, 30 June–3 July, Burlington, VT (1997).
- Gray, F. M. *Solid Polymer Electrolytes-Fundamentals and Technical Applications* (VCH, Weinheim, 1991).
- Evans, J., Vincent, C. A. & Bruce, P. G. Electrochemical measurements of transference numbers in polymer electrolytes. *Polymer* **28**, 2325–2330 (1987).
- Abraham, K. M., Jiang, Z. & Carroll, B. Highly conductive PEO-like polymer electrolytes. *Chem. Mater.* **9**, 1978–1988 (1997).
- Scrosati, B. Challenge of portable power. *Nature* **373**, 557–558 (1995).
- Croce, F. & Scrosati, B. Composite polymer ionics: advanced electrolyte materials for thin-film batteries. *Polym. Adv. Techn.* **4**, 198–204 (1993).

Acknowledgements. We thank M. Musei (CISE Laboratory, Milan) for providing samples of TiO_2 nanopowders.

Correspondence and requests for materials should be addressed to B.S. (e-mail: scrosati@axrma.uniroma1.it).

A synthetic mimic of the secretory granule for drug delivery

Patrick F. Kiser^{*†}, Glynn Wilson[‡] & David Needham^{*}

^{*} Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708-0300, USA

[†] Access Pharmaceuticals, Dallas, Texas 75207-2107, USA

[‡] The Glynn Wilson Group, 700 NW Gilman Boulevard E103, Issaquah, Washington 98027-5335, USA

Secretory cells contain submicroscopic granules composed of a polyanionic polymer network that is collapsed owing to the presence of hydronium ions and weak base cations^{1–3}. The network is encapsulated within a lipid membrane, and functions as a vehicle for the osmotically inert storage of a variety of granule-bound endogenous mediator species, such as histamine, serotonin and proteases. These species are excreted from the granule and thence from the cell in response to external biochemical signals^{1–4}. Hydrogels that swell and shrink in response to external stimuli might serve as synthetic analogues of secretory granules^{5,6}. Here we describe the systematic engineering of multi-component, environmentally responsive hydrogel microspheres, coated with a lipid bilayer to mimic more closely the natural secretory granule. These microspheres exhibit pH- and ion-dependent volume phase transitions and ion-sensitive exchange of bound cations when the encapsulating lipid membrane is porated. We stimulated poration electrically in individual microgel particles immobilized and manipulated with a micropipette. This system could find use for the triggered release of encapsulated drugs in the body.

The anionic microgels, composed of a 1:4 mole ratio of methylene-bis-acrylamide and methacrylic acid, were made by a method based on the precipitation polymerization developed by Kawaguchi^{7,8} and had a diameter of 6.5 μm at pH 7 (Fig. 1a and legend). They were loaded with doxorubicin (DX; ref. 6) by incubation of the negatively charged microgels with the cationic form of doxorubicin. Doxorubicin was chosen because it is a weak base cation and because it is a potent anticancer drug. Moreover, it is easily detected fluorimetrically and optically; thus, its concentration can be readily quantified in both the microgel and in solution. The maximum loading capacity of the microgel for DX at pH 7.0 was 4.31 ± 0.07 mmol DX per g polymer ($\pm$ s.d., $n = 4$), representing a DX concentration in the polymer microgel of >2 M. Comparatively, this loading capacity is ten times greater than the concentration of precipitated DX in the lumen of remote loaded liposomes (0.15–0.3 M; ref. 9) and is 100 times greater than DX in proteinaceous microspheres¹⁰, so offering a potentially greater delivery of this and similar substances for the same-sized particle.

Like the secretory granule, these microgels showed a strong volume dependence on pH (refs 2, 9). When the pH of the bathing medium was changed from 3.2 to 7.0, a volume change of $>300\%$ was observed. Essentially, an expanding Donnan pressure, due to deprotonation of the carboxyl groups, is balanced by the elasticity of the polymer network^{8,11}. The lumen of the secretory granule is acidified, thus, to reproduce this feature, we loaded the microgels at pH 5.0 in a solution of 1 mM DX in 5 mM citrate buffer, to give a DX-loading capacity in the microgel of ~ 1 M. In this state, the microgel was sufficiently condensed that significant expansion could occur on subsequent reswelling. The DX-loaded microgels immediately reswelled when transferred to phosphate-buffered saline (PBS; see Fig. 1a legend) at pH 7.4 due to dissipation of the pH gradient. The degree of swelling⁸ and the DX-loading capacity at

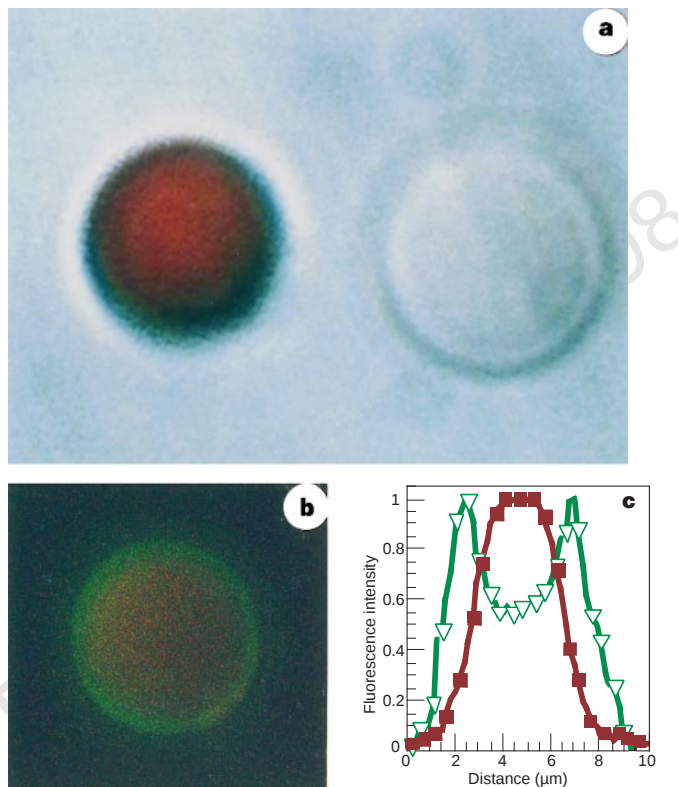


Figure 1 Videomicrographs of membrane-coated and uncoated microgels in PBS. **a**, Differential interference contrast video micrograph of a DX-loaded membrane bound microgel (dark red sphere, left) and a control unmodified microgel (translucent sphere, right); the drug loaded microgel appears deep red due to the presence of DX (λ_{max} , 540 nm); PBS was iso-osmotic with plasma (osmotic pressure, 270 mOsmol kg^{-1} ; $[\text{Na}^+] = 170$ mM; $[\text{PO}_4^{3-}] = 50$ mM). The lipid membrane stabilizes the condensed phase of the microgel by maintaining a hydrogen-ion gradient of 2.4 pH units in a solution which would otherwise swell the microgel and exchange DX. **b**, Dual-exposure fluorescence micrograph of a membrane-coated microgel. The green annulus is due to the membrane dye cholesteryl Bodipy-fl C₁₂ and the red emission is due to the native fluorescence of the DX anthraquinone fluorophore. **c**, Fluorescence intensity versus distance across the microgel centre of both the membrane dye tracer (green open triangles) and the DX (red filled squares) obtained from a membrane-coated microgel in **b**. The microgels are synthesized as a copolymer between methacrylic acid, 4-nitrophenyl methacrylate and *N,N'*-methylenebisacrylamide with a feed ratio of 2:2:1 (25 mmol total) in a 3:1 v/v mixture of ethanol and methanol (28 ml total). The free radical polymerization reaction was initiated by adding a solution of azobisisobutyronitrile (1.5 g) in degassed acetone (12 ml). The nitrophenyl groups are necessary for the formation of the particles and are subsequently removed by hydrolysis with hydroxide. The particles are purified by several centrifugation and wash cycles^{7,8}.

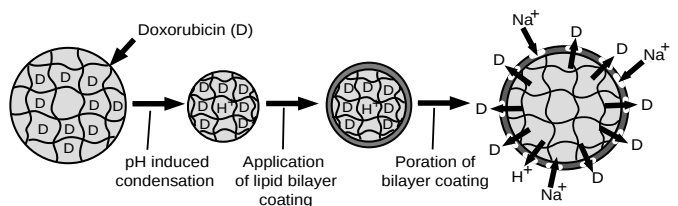


Figure 2 Diagram of the loading and release process for the synthetic mimic of the secretory granule. The microgel is loaded with doxorubicin (D) and condensed. By coating the condensed microgel with a lipid bilayer, the condensed state of the gel particle is stabilized in a solution which would otherwise swell the microgel. Poration of the bilayer causes the gel to swell, creating excess microgel surface area and allowing release doxorubicin via ion exchange.

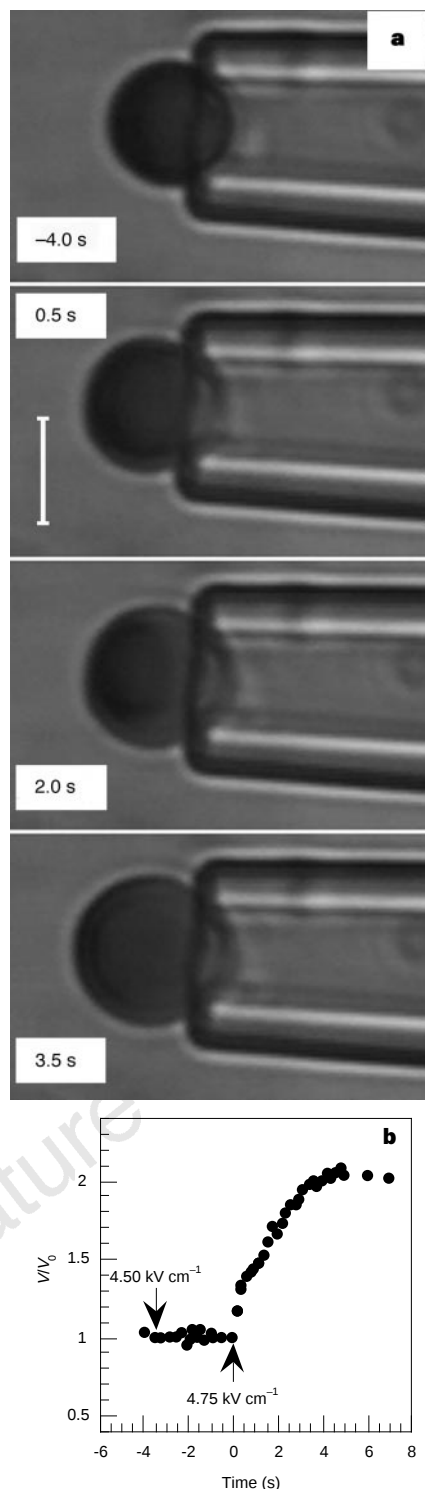


Figure 3 Swelling kinetics after electroporation of a microgel in PBS. **a**, Bright-field video micrographs showing a DX-loaded, membrane-coated and condensed microgel before and after electroporation in PBS; the microgel swells as the pH gradient across the bilayer is dissipated. The first image (labelled “-4.0 s”) was obtained 4 s before application of the electroporating pulse. The next image (labelled “0.5 s”) was collected 0.5 s after the application of the electroporating pulse (4.75 kV cm^{-1}) which caused membrane breakdown and allowed the microgel to swell. The next two images were obtained in 1.5-s intervals. Scale bar, $4 \mu\text{m}$. **b**, Time course of the microgel volume (V) compared to its value in the condensed state (V_0) obtained during sequential application of incrementally increasing electric fields ($t = 0$ indicating application of the pulse). At $t = -3.5$ s a field of 4.50 kV cm^{-1} was applied (arrow) which did not electromechanically disrupt the bilayer. At $t = 0$ s this field created an electropore and induced swelling (arrow). It took ~ 4 seconds for the microgel to reach its maximum volume.

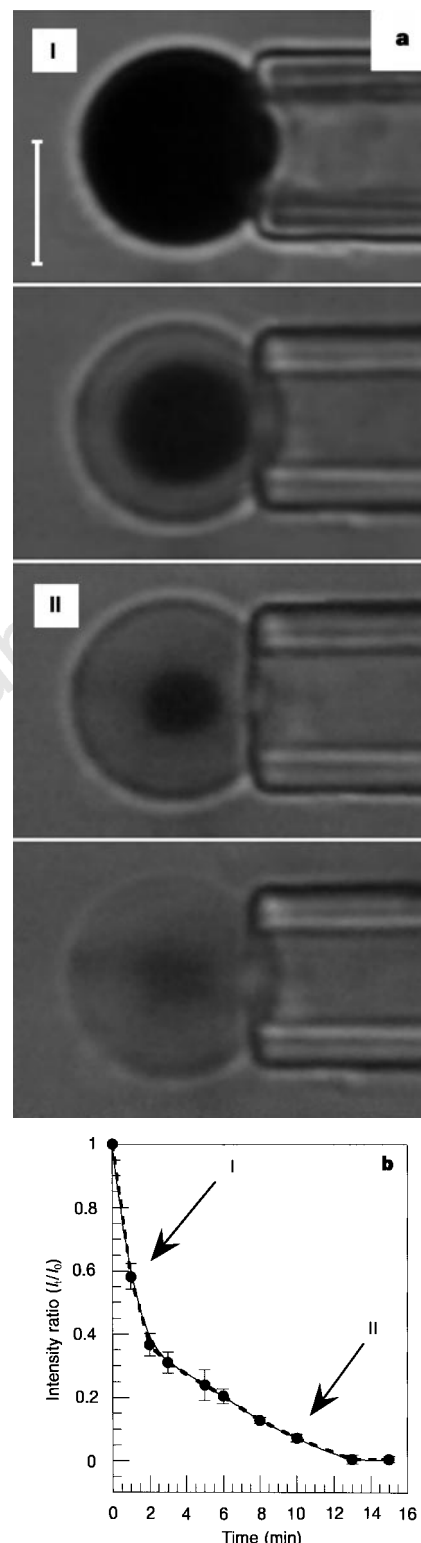


Figure 4 Release kinetics of doxorubicin from an electroporated microgel in PBS. **a**, A typical sequence of video micrographs showing release of DX from an electroporated microgel (images were collected through a green filter which caused the microgel to appear darker than those in Fig. 2a). The first image (labelled “I”) shows the swollen microgel 30 s after it had reached its maximum volume. The remaining images were obtained at 5, 10 (labelled “II”), and 13 min after this time. Scale bar, $4 \mu\text{m}$. **b**, Intensity ratio (I_t/I_0) versus time. The intensity ratio was obtained from image analysis of digitized video micrographs. The intensity values at t_i are normalized to the integrated grey-scale value I_0 obtained at the point at which the microgel had reached its maximum volume ($t = 4$ s in Fig. 2b). Labelled arrows I and II correspond to the labels on the video micrographs in Fig. 3a. The error bars represent the standard deviation of five separate measurements.

acidic pH were similar to those found for isolated secretory granules^{3,12}.

Secretory-cell stimulation causes the lipid membrane encapsulating the secretory granule and the plasma membrane of the cell to fuse, forming an aqueous pore between the cell exterior and the granule interior^{3,4,13}. This results in an ion-exchange-driven, diffusive release of its contents from the cell in a burst profile^{4,14}. Thus membrane-coated secretory granules do not swell and release their contents in physiological media unless their membranes are porated¹⁵. To prevent swelling and release of DX from our microgels in physiological media, the next stage of the assembly was to coat the microgel with a lipid bilayer. The pH condensed and DX-loaded microgels (see above) were incubated with hydrating lipid: dipalmitoylphosphatidylcholine, dipalmitoylphosphatidylglycerol sodium salt and cholesterol at a molar ratio of 4:1:5. Uncoated microgels were separated from membrane-coated microgels by their difference in density using gradient centrifugation.

Figure 1a shows a micrograph of a doxorubicin-loaded, membrane-coated microgel (deep red, left) in PBS, and an adjacent unmodified microgel (fully swollen, right) which was not exposed to the lipid coating process. To visualize the lipid bilayer coating, a lipophilic green fluorescent dye, cholesteryl Bodipy-fl C₁₂ (Molecular Probes, Eugene, Oregon) was included in the lipid-coating mixture. In Figure 1b we show a dual-colour fluorescence micrograph of the DX fluorescence (red) and the membrane dye cholesteryl Bodipy-fl C₁₂ (green). As shown in Fig. 1c, the red DX cross-section has maximum fluorescence intensity at the centre of the microgel, denoting bulk loading of the DX. The distribution of green lipid fluorescence shows a maximum at the edges of the microgel, indicating the presence of lipid only on the surface of the microgel and not in the interior. The fluorescence intensity value for the lipid bilayer coating was similar to single unilamellar lipid vesicles formed in the same suspension, indicating that the microgel coating was probably a single bilayer.

To permeate the bilayer coating in a controlled fashion and to show the critical nature of the swelling and DX release for individual microgels (Fig. 2), we developed a methodology using micropipette manipulation and electroporation^{15,16}. Figure 3 shows a sequence of video images of a DX-loaded and lipid-coated microgel in PBS before and after membrane permeation. A series of d.c. square-wave voltage pulses of increasing strength were applied to the microgel, and the lipid bilayer coating was electromechanically disrupted at a critical applied electric field strength of 4.75 kV cm⁻¹. This disruption of the protective bilayer coating allowed the condensed microgel to reach maximum swelling over a period of ~4 s as the pH gradient across the membrane ($\Delta\text{pH} = 2.4$) was dissipated (Fig. 3b). The corresponding increase in surface area of the microgel was 60%, leaving much of the microgel surface free of membrane and allowing diffusion of Na⁺ into, and DX and H⁺ out of, the microgel. In contrast to the unloaded microgel that swelled in ~300 ms (ref. 8) (approximately the same rate as the native secretory granule¹⁴), the observed swelling kinetics (Fig. 3b) were 20 times slower than those seen previously in electroporation experiments on intact mast cell granules in plasma saline¹⁵. This difference probably reflects the large amount of DX present in the network causing the exchange of Na⁺ to be rate limiting and moreover increasing the microviscosity within the gel network¹⁴. Both effects would decrease the swelling rate.

In re-creating the secretory granule structure, it was necessary to reproducibly achieve and demonstrate single-membrane coverage of the microgel. We therefore attempted to shear off multiple bilayers by filter extrusion¹⁷. As discussed above, the fluorescence data appeared to show only single membrane coverage. This was confirmed from the values of transmembrane potential required for electromechanical breakdown, which are related to the thickness of the dielectric layer surrounding the microgel. Mechanical data^{18,19} and electrocompression theory¹⁶ predict that an applied transmem-

brane potential of 3.2 V would be sufficient to induce electromechanical breakdown of a single lipid bilayer of the composition described above. The experimentally determined value was 3.5 ± 0.3 V ($\pm$ s.d., $n = 5$), so confirming the breakdown of a single bilayer¹⁹.

After bilayer disruption and microgel swelling, the DX-release kinetics were monitored for a single microgel by bright-field microscopy. Figure 4a shows a sequence of video micrographs (which follows swelling and electroporation shown in Fig. 3) in which DX is seen to radially diffuse out of the swollen microgel with increasing time, as the deep-red DX was exchanged for Na⁺ ions in the unstirred bathing solution. A translucent annulus formed around the periphery of the microgel, and ultimately the DX-loaded dark core dissipated, resulting in a translucent microgel (see Fig. 1a). The time course of the release was obtained by integrating the grey-scale values in these sequences to give a $t_{1/2}$ (time where the integrated intensity of the microgel was one-half that of the initial intensity) for the release of DX of 2 min (Fig. 4b). By comparison, the $t_{1/2}$ for the release of serotonin from native granules is ~150 ms (ref. 15); that is, three orders of magnitude faster than the microgels of similar size. These differences are not surprising for the concentration of serotonin in the granule matrix (~3–20 mM; ref. 14) is 100 times smaller than that of DX in the microgel (~1 M). The presence of high concentrations of small molecules in native granules has already been shown to reduce the diffusivity of the gel network and exchangeable ions within the gel, thus reducing the swelling and release rates¹⁴.

Our mimic of the secretory granule thus provides a controllable triggered burst release of a drug—release occurs only when the permeability barrier of the protecting membrane is broken down. This could be done, for example, by the *in vivo* application of external heating to temperature-sensitive bilayers or by poration induced by ultrasound. Transmembrane channels that are triggered to open by an external signal and thus permeate the membrane are also being considered. This rapid, triggerable release feature would not occur for macroscopic slab gels that swell and exchange material over periods of hours to days (ref. 20). This approach could also be extended to provide gel particles in the nanometre size range that, in principle, are small enough to pass through permeable tumour vasculature into tumour tissue²¹ and, ultimately, to be taken up by cells. □

Received 28 January; accepted 13 May 1998.

1. Metcalf, D. D., Kaliner, M. & Donlon, M. A. The mast cell. *Crit. Rev. Immunol.* **3**, 23–74 (1981).
2. Curran, M. J. & Brodwick, M. S. Ionic control of the size of the vesicle matrix of beige mouse mast cells. *J. Gen. Physiol.* **98**, 771–790 (1991).
3. Fernandez, J. M., Villalon, M. & Verdugo, P. Reversible condensation of mast cell secretory products in vitro. *Biophys. J.* **59**, 1022–1027 (1991).
4. Nanavati, C. & Fernandez, J. M. The secretory granule matrix: a fast acting smart polymer. *Science* **258**, 963–965 (1993).
5. Fernandez, J. M. & Knutson, M. B. *Condensed-Phase Microparticle Composition and Method* US Patent No. 5,654,006 (1997).
6. Kiser, P. F., Needham, D. & Wilson, G. A synthetic mimetic of the secretory granule: a new concept in controlled release drug delivery. *Proc. Am. Chem. Soc. Div. Polym. Mater. Sci. Eng.* **76**, 226 (1997).
7. Kashiwabara, M., Fujimoto, K. & Kawaguchi, H. Preparation of monodisperse, reactive hydrogel microspheres and their amphotericization. *Colloid. Polym. Sci.* **273**, 339–345 (1995).
8. Eichenbaum, G., Kiser, P. F., Simon, S. & Needham, D. pH and ion triggered volume response of anionic hydrogel microspheres. *Macromolecules* (in the press).
9. Madden, T. D. et al. The accumulation of drugs within large unilamellar vesicles exhibiting a proton gradient: a survey. *Chem. Phys. Lipids* **53**, 37–46 (1990).
10. Chen, Y., Burton, M. A. & Gray, B. N. in *Microspheres and Regional Cancer Therapy* (eds Wilmont, N. & Daly, J.) 2–23 (CRC, Boca Raton, FL, 1994).
11. Ricka, J. & Tanaka, T. Swelling of ionic gels: quantitative performance of the donnan theory. *Macromolecules* **17**, 2916–2921 (1984).
12. Verdugo, P., Orellana, M. & Freitag, C. The secretory granule as a biomimetic model for drug delivery. *Proc. Int. Symp. Controlled Release Bioact. Mater.* **114** (1995).
13. Alvarez de Toledo, G., Fernandez-Chacon, R. & Fernandez, J. M. Release of secretory products during transient vesicle fusion. *Nature* **363**, 554–557 (1993).
14. Marszalek, P. E., Farrell, B., Verdugo, P. & Fernandez, J. M. Kinetics of release of serotonin from isolated secretory granules. (ii) ion exchange determines the diffusivity of serotonin. *Biophys. J.* **37**, 1169–1183 (1997).
15. Marszalek, P. E., Farrell, B., Verdugo, P. & Fernandez, J. M. Kinetics of release of serotonin from isolated secretory granules. (i) amperometric detection of serotonin from electroporated granules. *Biophys. J.* **37**, 1160–1168 (1997).
16. Needham, D. & Hochmuth, R. M. Electro-mechanical permeabilization of lipid vesicles: role of membrane tension and compressibility. *Biophys. J.* **55**, 1001–1009 (1989).
17. Mayer, L. D., Hope, M. J. & Cullis, P. R. Vesicles of variable sizes produced by rapid extrusion procedure. *Biochim. Biophys. Acta* **858**, 161–168 (1986).

18. Needham, D. & Nunn, R. S. Elastic deformation and failure of lipid bilayer membranes containing cholesterol. *Biophys. J.* **58**, 997–1009 (1990).
19. Needham, D. & Zhelev, D. in *Vesicles* (ed. Rosoff, M.) 374–444 (Marcell Dekker, Inc., New York, 1996).
20. Jeong, J., Bae, Y. H., Lee, D. S. & Kim, S. W. Biodegradable block copolymers as injectable drug-delivery systems. *Nature* **388**, 860–862 (1997).
21. Wu, N. Z. *et al.* Increased microvascular permeability contributes to preferential accumulation of Stealth liposomes in tumor tissue. *Canc. Res.* **53**, 3765–3770 (1993).

Acknowledgements. We thank J. M. Fernandez for discussions, and P. Marszalek for guidance on microelectroporation. We also thank Access Pharmaceuticals Inc., Dallas, Texas for financial support.

Correspondence and requests for materials should be addressed to D.N. (e-mail: david.needham@duke.edu).

Budgetary and biogeochemical implications of N₂O isotope signatures in the Arabian Sea

S. W. A. Naqvi*, T. Yoshinari†, D. A. Jayakumar*, M. A. Altabet‡, P. V. Narvekar*, A. H. Devol§, J. A. Brandes§|| & L. A. Codispoti¶

* National Institute of Oceanography, Dona Paula, Goa 403 004, India

† Wadsworth Center, New York State Department of Health, and School of Public Health, State University of New York at Albany, PO Box 509, Albany, New York 12201, USA

‡ Department of Chemistry and Biochemistry, and Center for Marine Science and Technology, University of Massachusetts, Dartmouth, North Dartmouth, Massachusetts 02747, USA

§ School of Oceanography, University of Washington, Seattle, Washington 98195, USA

¶ Center for Coastal Physical Oceanography, Old Dominion University, Norfolk, Virginia 23529, USA

Nitrous oxide (N₂O) is an important greenhouse gas that also plays a role in the chemistry of stratospheric ozone depletion, but its atmospheric budget has yet to be well-quantified^{1–5}. However, multi-isotope characterization of N₂O emitted from various natural sources is a potentially powerful tool for providing the much-needed constraints. It is generally believed that production of isotopically light (low ¹⁵N/¹⁴N and ¹⁸O/¹⁶O ratios) N₂O occurs in the upper ocean through nitrification process, and that the flux of this light N₂O from sea to air isotopically counters the flux of heavy N₂O from the stratosphere to the troposphere^{1,2}. But eastern-boundary ocean-upwelling zones, which contain oxygen-depleted waters and are sites of intense N₂O efflux^{6–10}, have not been adequately studied. We show here, using new isotope data, that in spite of huge denitrification-related enrichments of ¹⁵N and ¹⁸O in N₂O at mid-depths in the Arabian Sea, N₂O emitted from upwelled waters is only slightly enriched in ¹⁸O, and moderately depleted in ¹⁵N, relative to air. These opposing isotopic signatures and modest departures from the isotopic composition of tropospheric N₂O indicate that air–sea exchange cannot—given the heavy isotopic signature of N₂O derived from the stratosphere—allow the tropospheric budget of N₂O to be closed without invoking hitherto-unknown N₂O sources and sinks. Our oceanic data cannot be explained through either nitrification or denitrification alone, such that a coupling between the two processes may be an important mechanism of N₂O production.

Estimates of the rate of N₂O emission from the ocean to the atmosphere are currently being revised upwards as a result of acquisition of new data from coastal environments, especially those affected by upwelling. The most recent estimate⁹ of oceanic source (7–11 Tg N yr^{–1}; 1 Tg = 10¹² g) compares favourably with the biologically constrained figure (11 Tg N yr^{–1}; ref. 11). As these values approach the estimated stratospheric loss rate (12.5 Tg N yr^{–1}; ref. 2), they probably represent an upper limit, but they highlight the

role of the ocean in regulating the level, and isotopic composition of, atmospheric N₂O.

Recent measurements of N₂O in the upwelling zones off Oman¹⁰ and Somalia¹² revealed large surface saturations (up to 230% and 330%, respectively), indicating that the annual flux of N₂O from the Arabian Sea alone may be as much as 0.5–1 Tg N yr^{–1}. Our observations in the lesser-known third upwelling centre (off southwestern India) yielded even more extreme data: at 23 stations (latitudes 8.58–11.85° N, longitudes 74.34–76.32° E) sampled during the southwest monsoon of 1995, surface concentrations and saturations were 11.2–62.5 nM (mean ± s.d. = 28.5 ± 14.7 nM) and 193–953% (458 ± 223%), respectively. These values, among the highest observed in oceanic surface waters⁸, were associated with low sea-surface temperatures (minimum 22.8 °C) and high NO₃[–] concentrations (maximum 16 μM). The ¹⁵N/¹⁴N and ¹⁸O/¹⁶O ratios in surface-water N₂O (expressed as δ¹⁵N and δ¹⁸O relative to air) at eight of these sites, and at six others in the northern Arabian Sea, are plotted versus percentage saturation in Fig. 1. The δ¹⁵N and percentage saturation appear to be inversely related, with N₂O in upwelled waters off southwestern India being the most depleted in ¹⁵N. A δ¹⁵N value of 0.8‰, the lowest ever reported for sea water, was recorded at 26 m depth at 11.85° N, 75.15° E; the corresponding N₂O concentration was 115.6 nM. Because at such a high concentration the isotopic signature of background N₂O would be erased, this value can be taken to represent the lightest end-member composition of N₂O produced in sea water. It is substantially lower than the δ¹⁵N of tropospheric N₂O (7‰; ref. 13), but much higher than expected from the reported depletion of ¹⁵N (less than –60‰ relative to NH₄⁺) during nitrification¹⁴. In contrast to δ¹⁵N, δ¹⁸O did not exhibit a discernible relationship with percentage saturation, and with the exception of one sample, all δ¹⁸O values were consistently higher than the δ¹⁸O of tropospheric N₂O (20.7‰; ref. 13).

The δ¹⁵N of surface-water N₂O at the six open ocean locations is close to the tropospheric value (Fig. 1), taking into account the 0.8‰ enrichment in the aqueous phase¹⁵. To determine how surface-water isotopic composition is influenced by processes within

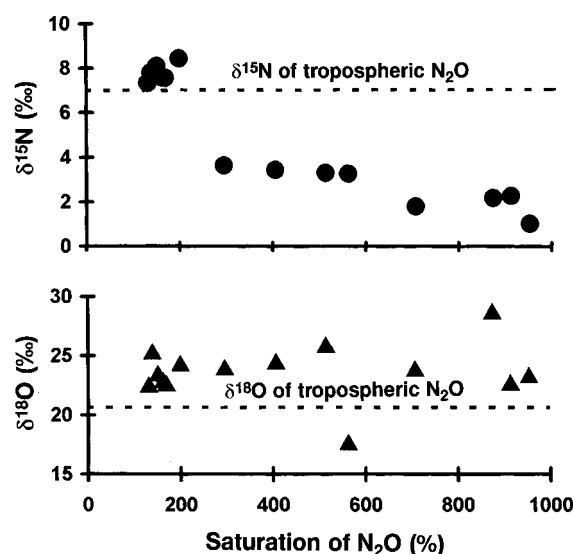


Figure 1 Natural isotope ratios (δ¹⁵N and δ¹⁸O relative to air) in dissolved N₂O as a function of N₂O saturation at the sea surface (depth < 5 m). Horizontal dashed lines indicate the isotopic composition of N₂O in the troposphere¹³. Values with moderate saturation and δ¹⁵N close to the tropospheric value are from the northern Arabian Sea (latitudes 15.67–20.03° N, longitudes 61.02–68.00° E). Two of these sites were sampled aboard RV *Sagar Sampada* in April 1994, while the other four were sampled aboard RV *Sagar Kanya* in July 1995. All other data were taken from the coastal upwelling zone of southwestern India (latitudes 8.58–11.85° N, longitudes 74.34–76.32° E) during the latter cruise.

|| Present address: Geophysical Laboratory, Carnegie Institution of Washington, Washington DC 20015, USA.

the underlying suboxic layer, we examined vertical isotope profiles (Fig. 2) at two stations of the RV *Sagar Sampada*, one of which (SS 3201) was located within the zone of intense denitrification while the other (SS 3204) was situated at its periphery¹⁶. The N_2O minimum observed at 200–500 m depth at SS 3201 coincided with an intense secondary nitrite maximum and a pronounced peak of $\delta^{15}\text{N}$ of NO_3^- (Fig. 2). Preferential loss of lighter N_2O to N_2 within this layer evidently led to enrichments of ^{15}N and ^{18}O in residual N_2O which are by far the greatest reported from any natural environment^{1,4,15,17,18}. The N_2O concentration minimum was embedded between two maxima, but the isotopic compositions of these N_2O -rich layers were very different. Going upwards from the oxygen minimum zone the $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ values fell sharply across the oxic–suboxic boundary with the former generally lower than the tropospheric value in the upper 150 m (Fig. 2). In contrast, although $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ also decreased below the secondary nitrite maximum, levels remained much higher than the tropospheric values down to 1,500 m. A similar pattern was also seen at the mildly reducing SS 3204, although the $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ increases within the oxygen minimum zone were smaller. The differences between redox conditions at the two stations probably resulted from slightly different O_2 levels that were low enough to be masked by analytical errors associated with the Winkler titrations (Fig. 2). Because the suboxic waters with large enrichments of heavier isotopes are capped by a layer which has high concentrations of light N_2O , the N_2O escaping to the atmosphere in the region is not heavy. Thus the near-tropospheric $\delta^{15}\text{N}$ values confined to the shallowest depths at the open ocean sites (Figs 1 and 2) probably reflect an active air–sea exchange. The fate of heavy N_2O is not known, but it is expected to be advected out of the region and to contribute to the elevated $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ in N_2O observed outside the suboxic zones¹³.

The observed trends in $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ in N_2O -rich surface waters of the Arabian Sea have important implications for the atmospheric N_2O budget. Isotopic analysis of stratospheric N_2O at 68° N has shown large enrichments of $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ relative to the troposphere¹. This trend has been confirmed⁴ by more recent observations at 33–48° N, albeit with somewhat smaller enrichments in the lower stratosphere, but the causative mechanism is still not fully understood^{2–5}. However, the magnitude of ^{15}N and ^{18}O fluxes (~400 and 500 Tg N % yr^{-1} , respectively, obtained from the older data¹ and about half as much from recent results⁴) are such that these cannot be balanced by air–sea exchange, as proposed^{1,2}. Even if, as an extreme case, we combine the lowest observed $\delta^{15}\text{N}$ value from the ocean (~1‰) with the largest estimate of oceanic N_2O production (11 Tg N yr^{-1} ; ref. 11), the air–sea exchange can counter no more than 66 Tg N % yr^{-1} of the stratospheric ^{15}N flux. The actual oceanic N_2O production is probably smaller, and its average $\delta^{15}\text{N}$ higher than the above values. If, for example, we

assume that half of the N_2O production is through denitrification¹¹ (with characteristically much higher $\delta^{15}\text{N}$ —see below), the oceanic contribution to neutralize the stratospheric ^{15}N enrichment will be <30 Tg N % yr^{-1} . The reported ^{18}O fluxes are even more problematic, as the opposite isotopic trends observed by us imply that the air–sea exchange will lead to an increase in the tropospheric ^{18}O inventory. The importance of this issue is evident from the fact that on a $\delta^{15}\text{N}$ versus $\delta^{18}\text{O}$ diagram constructed with the published data representing various planetary reservoirs of N_2O , the stratospheric composition conspicuously forms a ‘corner’ (Fig. 3). Thus the tropospheric value, which does not fall on vectors joining the other end-members, cannot be explained by simple mixing between various reservoirs¹.

It may be argued that our data may not be representative of the oceans as a whole, and that sources of lighter N_2O could exist elsewhere (for example, off Peru, where large N_2O supersaturations have also been found at the sea surface)⁸. However, this possibility is very remote. The Arabian Sea contains several diverse biogeochemical environments and offers the most extreme conditions for N_2O cycling¹⁹. Consequently, some of the highest and lowest N_2O concentrations in the world oceans are found here⁷; similarly, the isotope ratios are also the most extreme reported yet from the oceans^{8,13,15,17,18}. These extreme isotope ratios may therefore be used to constrain the maximum contribution that the oceans can make to counter heavy-isotope enrichment in stratospheric N_2O . In view of these observations, more measurements need to be made on N_2O in the stratosphere and that emitted from different terrestrial reservoirs. If the published data are representative, however, then there ought to exist some hitherto poorly known sources and/or sinks^{2,3,20} of N_2O that may be vital for tropospheric isotopic balance.

Our results also provide some insights into the mechanisms of N_2O production. An important aspect of the data is that at both the open-ocean sites $\delta^{15}\text{N}$ of NO_3^- is consistently lower than that of N_2O at depths exceeding ~200 m, and higher at shallower depths (Fig. 2). This indicates that the intense N_2O accumulation in the upper and lower parts of the oxygen minimum zone may be through different mechanisms. Greater enrichment of ^{15}N in N_2O relative to NO_3^- appears to be characteristic of denitrification, as revealed by recent culture experiments (M.A.A., unpublished results). This pattern is the most pronounced within the N_2O -depleted secondary nitrite maximum at SS 3201. Its persistence even below this maximum at SS 3201, and at depths exceeding ~200 m at SS 3204, suggests significant N_2O production through denitrification. This is because the concentrations of heavy N_2O within the secondary nitrite maximum are too low for mixing to fully account for the observed enrichment of heavier isotopes outside this layer. This observation is consistent with the view that denitrification may lead to a net N_2O accumulation under certain conditions^{11,21}.

Although the isotopically lighter N_2O in waters above the suboxic

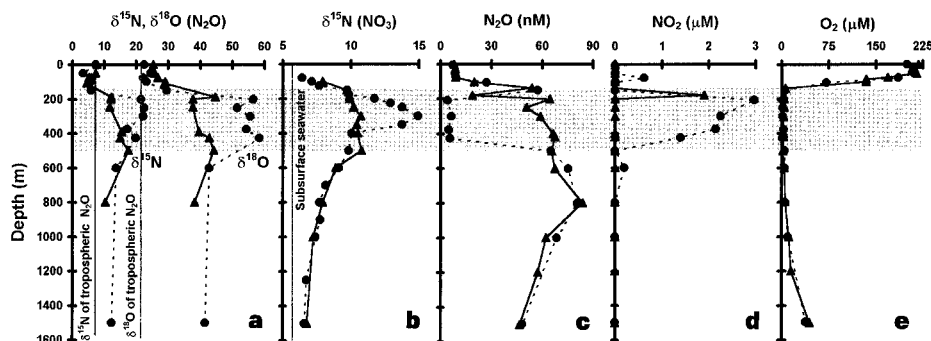


Figure 2 Vertical profiles of $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ (‰ relative to air) of N_2O (a), $\delta^{15}\text{N}$ of NO_3^- (b), and concentrations of N_2O (c), NO_2^- (d) and O_2 (e) at stations SS 3201 (17° N, 68° E; filled circles, joined by dotted lines) and SS 3204 (19.75° N, 64.62° E; filled triangles, joined by continuous lines). The shaded region represents the second-

ary nitrite maximum at SS 3201. Vertical lines indicate $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ of tropospheric N_2O (a) and mean $\delta^{15}\text{N}$ of NO_3^- in subsurface sea water²⁷ (b). Both stations were occupied in April 1994. The $\delta^{15}\text{N}$ (NO_3^-) profile at the SS 3201 site was obtained in January 1995.

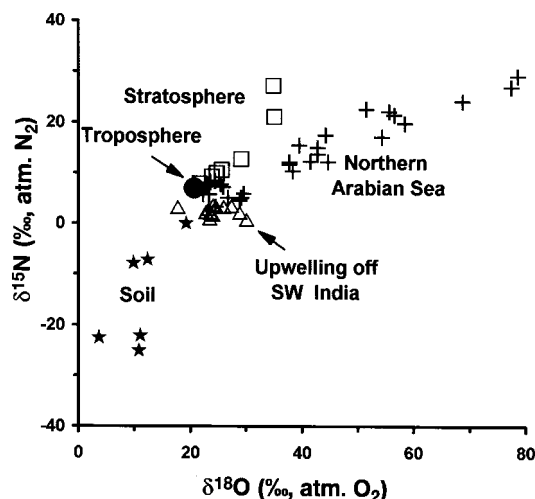


Figure 3 Comparison of natural isotope data from this study with representative data from other environments¹⁴. We note that the tropospheric value does not fall on vectors joining the other end-members.

zone is apparently not produced by denitrification, the classical nitrification pathway ($\text{NH}_4^+ \rightarrow \text{NO}_2^- \rightarrow \text{N}_2\text{O}$; ref. 14) also cannot fully explain the observed data for two reasons. First, as stated earlier, the $\delta^{15}\text{N}$ values are much higher than that reported for the production of N_2O from NH_4^+ by *Nitrosomonas europaea*¹⁴. Second, what is more significant is that the observed $\delta^{18}\text{O}$ is much higher than the value for nitrification (less than -3‰) expected¹⁸ from the $\delta^{18}\text{O}$ of H_2O and O_2 . A more likely pathway of N_2O production in the upper layer, which can account for both the isotope and concentration data, is a nitrification–denitrification coupling^{22,23} where an intermediate or by-product of nitrification such as NO may enter the denitrification sequence and get reduced to N_2O . As the oxygen atom of hydroxylamine (NH_2OH), the probable precursor²⁴ of NO , is derived²⁵ from O_2 which is characterized²⁶ by high $\delta^{18}\text{O}$, N_2O so produced is expected to be ^{18}O -enriched. Moreover, unlike N_2O , production of NO by nitrifying bacteria¹⁴ involves modest depletion of ^{15}N . Thus it may be concluded that there are several possible modes of N_2O production in the ocean involving different isotopic fractionations. The relative contribution from various pathways is probably determined by oxygen distribution and organic-carbon availability, both of which vary in space and time. This not only brings about large spatial changes in N_2O production and consumption but, in response to temporal changes in suboxic zones²⁷, the oceanic N_2O source strength may also vary greatly with time, contributing to fluctuations in atmospheric N_2O as recorded in ice cores²⁸.

Methods

N_2O dissolved in water taken with modified 20-l Go-flo bottles, mounted on a conductivity-temperature-depth (CTD) rosette-sampler, was extracted immediately after collection without transferring the samples to another vessel, and concentrated in stainless-steel tubes filled with molecular sieve (MS 5A). After thermal desorption in the on-shore laboratory, N_2O was purified by gas chromatography and its isotopic composition measured at Woods Hole Oceanographic Institution by continuous-flow isotope-ratio monitoring using a Finnigan MAT 251 mass spectrometer¹⁸. Analytical precision was better than 0.3‰ for both isotopes. NO_3^- isotope analysis was done at the University of Washington on filtered, HgCl_2 -poisoned samples using a Devarda's alloy reduction technique²⁹. N_2O concentration was determined on-board ship by head-space extraction with helium followed by injection into a Perkin-Elmer 3920B gas chromatograph equipped with an electron capture detector (precision $\sim 4\%$). Percentage saturation was computed with reference to the solubilities given in ref. 30. O_2 was estimated with the Winkler method, and nutrients were analysed using an autoanalyser⁷.

Received 28 July 1997; accepted 29 April 1998.

- Kim, K.-R. & Craig, H. Nitrogen-15 and oxygen-18 characteristics of nitrous oxide: a global perspective. *Science* **262**, 1855–1857 (1993).
- McElroy, M. B. & Jones, D. B. Evidence for an additional source of atmospheric N_2O . *Glob. Biogeochem. Cycles* **10**, 651–659 (1996).
- Cliff, S. S. & Thiemens, M. H. The $^{18}\text{O}/^{16}\text{O}$ and $^{17}\text{O}/^{16}\text{O}$ ratios in atmospheric nitrous oxide: a mass-dependent anomaly. *Science* **278**, 1774–1776 (1997).
- Rahn, T. & Wahlen, M. Stable isotope enrichment in stratospheric nitrous oxide. *Science* **278**, 1776–1778 (1997).
- Yung, Y. L. & Miller, C. E. Isotopic fractionation of stratospheric nitrous oxide: is there a chemical source? *Science* **278**, 1778–1780 (1997).
- Codispoti, L. A. & Christensen, J. P. Nitrification, denitrification and nitrous oxide cycling in the eastern tropical South Pacific Ocean. *Mar. Chem.* **16**, 277–300 (1985).
- Naqvi, S. W. A. & Noronha, R. J. Nitrous oxide in the Arabian Sea. *Deep-Sea Res.* **38**, 871–890 (1991).
- Codispoti, L. A. et al. in *Oceanography of the Indian Ocean* (ed. Desai, B. N.) 271–284 (Oxford-IBH, New Delhi, 1992).
- Bange, H. W., Rapsomanikis, S. & Andreae, M. O. Nitrous oxide in coastal waters. *Glob. Biogeochem. Cycles* **10**, 197–207 (1996).
- Bange, H. W., Rapsomanikis, S. & Andreae, M. O. Nitrous oxide emissions from the Arabian Sea. *Geophys. Res. Lett.* **23**, 3175–3178 (1996).
- Capone, D. G. A biologically constrained estimate of oceanic N_2O flux. *Mitt. Internat. Verein. Limnol.* **25**, 105–114 (1996).
- de Wilde, H. P. J. & Helder, W. Nitrous oxide in the Somali Basin: the role of upwelling. *Deep-Sea Res.* **44**, 1319–1340 (1997).
- Kim, K.-R. & Craig, H. Two isotope characterisation of N_2O in the Pacific Ocean and constraints on its origin in deep water. *Nature* **347**, 58–61 (1990).
- Yoshida, N. ^{15}N -depleted N_2O as a product of nitrification. *Nature* **335**, 528–529 (1988).
- Yoshida, N., Hattori, A., Saino, T., Matsuo, S. & Wada, E. $^{15}\text{N}/^{14}\text{N}$ ratio of dissolved N_2O in the eastern tropical Pacific Ocean. *Nature* **307**, 442–444 (1984).
- Naqvi, W. A. Geographical extent of denitrification in the Arabian Sea in relation to some physical processes. *Oceanol. Acta* **14**, 281–290 (1991).
- Yoshida, N. et al. Nitrification rates and ^{15}N abundances of N_2O and NO_3^- in the western North Pacific. *Nature* **342**, 895–897 (1989).
- Yoshinari, T. et al. Nitrogen and oxygen isotopic composition of N_2O from suboxic waters of the eastern tropical North Pacific and the Arabian Sea—measurement by continuous flow isotope-ratio monitoring. *Mar. Chem.* **56**, 253–264 (1997).
- Burkill, P. H., Mantoura, R. F. C. & Owens, N. J. P. Biogeochemical cycling in the northwestern Indian Ocean: a brief overview. *Deep-Sea Res.* **40**, 643–649 (1993).
- Prasad, S. S. Natural atmospheric sources and sinks of nitrous oxide. I. An evaluation based on ten laboratory data. *J. Geophys. Res.* **99**, 5285–5294 (1994).
- Betlach, M. R. & Tiedje, J. M. Kinetic explanation for accumulation of nitrite, nitric oxide and nitrous oxide during bacterial denitrification. *Appl. Environ. Microbiol.* **42**, 1074–1084 (1981).
- Naqvi, S. W. A. N_2O production in the ocean. *Nature* **349**, 373–374 (1991).
- Zafriou, O. C. Laughing gas from leaky pipes. *Nature* **347**, 15–16 (1990).
- Ward, B. B. & Zafriou, O. C. Nitrification and nitric oxide in the oxygen minimum of the eastern tropical North Pacific. *Deep-Sea Res.* **35**, 1127–1142 (1988).
- Dua, R. D., Bhandari, B. & Nicholas, D. J. D. Stable isotope studies in the oxidation of ammonia to hydroxylamine by *Nitrosomonas europaea*. *FEBS Lett.* **106**, 401–404 (1979).
- Ostlund, H. G., Craig, H., Broecker, W. S. & Spencer, D. *GEOSECS Atlantic, Pacific and Indian Ocean Expeditions Vol. 7, Shorebased Data and Graphics* (NSF, Washington, DC, 1987).
- Altabet, M. A., Francois, R., Murray, D. W. & Prell, W. L. Climate-related variations in denitrification in the Arabian Sea from sediment $^{15}\text{N}/^{14}\text{N}$ ratios. *Nature* **373**, 506–509 (1995).
- Leuenberger, M. & Siegenthaler, U. Ice-age atmospheric concentration of nitrous oxide from an Antarctic ice core. *Nature* **360**, 449–451 (1992).
- Brandes, J. A. & Devol, A. H. Isotopic fractionation of oxygen and nitrogen in coastal marine sediments. *Geochim. Cosmochim. Acta* **61**, 1793–1802 (1997).
- Weiss, R. F. & Price, B. A. Nitrous oxide solubility in water and seawater. *Mar. Chem.* **8**, 347–359 (1980).

Acknowledgements. We thank E. Desai and B.J. Zaharanec for their encouragement and support; the Department of Ocean Development (India), especially V. Ravindranathan, for making time on the research vessels available; and C. Nasnolkar, V.V.S.S. Sarma and M. Kuhlend for their technical assistance. This work was supported by the Office of Naval Research (USA) through the US–India Fund.

Correspondence and requests for materials should be addressed to S.W.A.N. (e-mail: naqvi@csnio.ren.nic.in).

Orbital modulation of the Earth's magnetic field intensity

J. E. T. Channell*†, D. A. Hodell*, J. McManus‡ & B. Lehman†

* Department of Geology, University of Florida, Gainesville, Florida 32611-7340, USA

† Laboratoire des Sciences du Climat et de l'Environnement, CNRS-CEA, 91198 Gif-sur-Yvette, France

‡ Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

More than 20 years ago, on the basis of data from a Pacific sediment core, it was suggested that geomagnetic field intensity may vary with the Earth's orbital obliquity (centred on a period of ~ 41 kyr) as a result of the effect of obliquity on precessional forces in the Earth's core¹. It had also been proposed that precession plays an important role in the energy budget of the Earth's

geodynamo². But subsequent analyses indicated that the energy available from precession is at least an order of magnitude less than that required to drive the geodynamo³. Here, however, we report a spectral analysis of sedimentary records of relative geomagnetic palaeointensity from two North Atlantic sites which shows significant power both at orbital eccentricity (~100 kyr) and obliquity (41 kyr). The eccentricity power is also present in bulk magnetic properties (such as susceptibility) and is therefore attributable to lithological variations controlled by eccentricity-driven climate change. The obliquity power, however, is not apparent in bulk magnetic properties, and seems to be a property of the geomagnetic field itself, thus providing evidence for the orbital forcing of geomagnetic field intensity.

Recovery by the Ocean Drilling Program of continuous sedimentary records with high accumulation rates allows the measurement of long, detailed, proxy records of relative palaeointensity that can be dated using oxygen-isotope stratigraphy. Spectral analysis can be used to test whether the intensity of the Earth's magnetic field has a periodic component. For example, the Mediterranean–Indian Ocean palaeointensity stack⁴ for the past 140 kyr showed spectral power at orbital frequencies; however, the short duration of the time series and the unknown influence of orbitally driven lithological change precluded firm conclusions as to the origin of the observed periodicities. Other records from the Ontong–Java plateau⁵ and the South Atlantic^{6,7} indicated significant spectral power in the 30–

50 kyr period range that was not seen in the bulk magnetic properties; however, it was concluded that the observed cyclicity was distinct in frequency from either orbital precession or obliquity.

Ocean Drilling Project (ODP) Site 983 (60.4° N, 23.6° W) and Site 984 (61.4° N, 24.1° W) are located south of Iceland on the Gardar and Bjorn drifts, respectively^{8,9}. At Site 983 and Site 984, the Brunhes–Matuyama (B/M) boundary lies at 81 and 98 metres below sea floor (m.b.s.f.), indicating mean sedimentation rates within the Brunhes chron of 10.4 cm kyr⁻¹ and 12.6 cm kyr⁻¹, respectively. A composite section was constructed for each site by correlating core logging data for three holes at each site, to produce a spliced record representing a complete and undisturbed record of the sedimentary sequence¹⁰. The tandem measurement of sediment magnetic properties and stable-isotope ratios of foraminifers in the composite section of sites 983 (Fig. 1) and 984 (Fig. 2) allowed the development of continuous proxy signals of geomagnetic palaeointensity (Figs 1c and 2c) that are tied to oxygen-isotope stratigraphies (Figs 1a and 2a).

Proxies for relative geomagnetic palaeointensity in sediments can be constructed by normalizing the natural remanent magnetization (NRM) by a laboratory-produced remanence, either isothermal remanent magnetization (IRM) or anhysteretic remanent magnetization (ARM). The ideal normalizer compensates for variations in concentration of remanence-carrying grains; therefore, the normalizer should activate the same grains as those that carry NRM. For

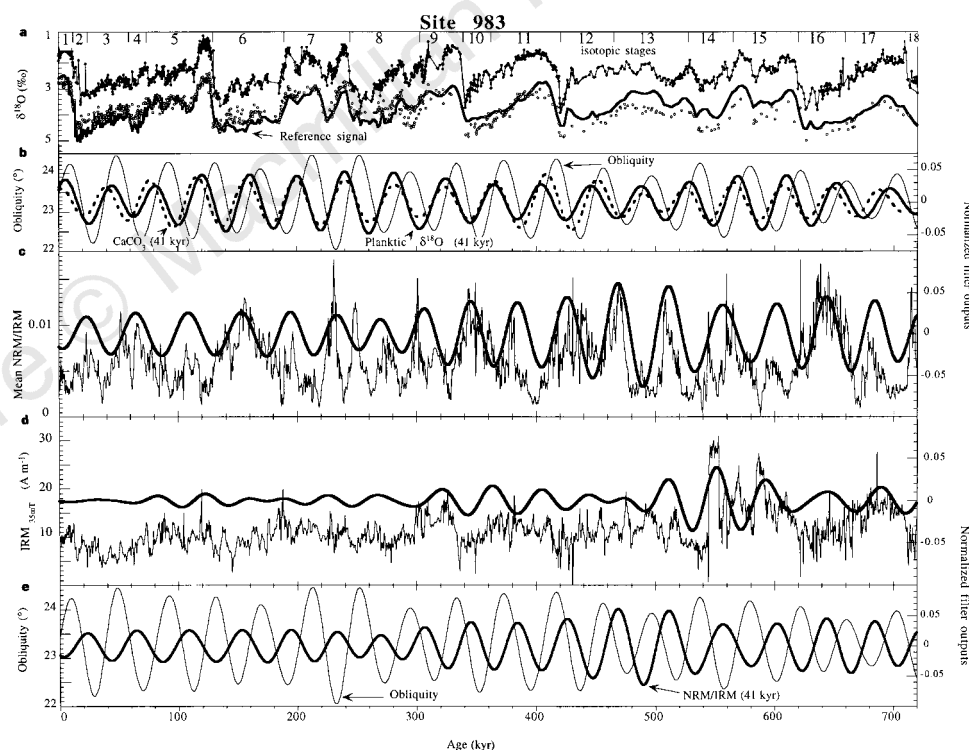


Figure 1 Site 983: relative geomagnetic palaeointensity proxy (mean NRM/IRM), the palaeointensity normalizer (IRM), planktic and benthic oxygen isotopes and filtered percentage CaCO_3 . **a**, Planktic (symbols joined by thin line) and benthic (symbols without line) oxygen-isotope ($\delta^{18}\text{O}$) records. Age models were constructed by correlation of the oxygen-isotope records to a standard reference signal (thick line) derived by combining the record of Core SU90-08 from 0 to 130 kyr ago²⁷ with the SPECMAP records from 130–290 kyr (ref. 22) and 290–350 kyr (ref. 23), with the ODP Site 677 record for 350–750 kyr (ref. 26). Linear interpolation between 35 tie points (to isotopic stage 18) produced the age model which yields an extrapolated age of 779 kyr for the Brunhes–Matuyama boundary. The planktic oxygen-isotope record was determined using tests of *Globigerina bulloides* selected from the 212–300 μm size fraction. The benthic oxygen isotopic record was constructed by measurement of specimens of *Cibicidoides wueller-*

storfi selected from the >150 μm size fraction. Additional specimens of *Cibicidoides* and *Uvigerina* were selected at intervals that were barren of *C. wuellerstorfi* and corrected to *C. wuellerstorfi* values. **b**, Planktic $\delta^{18}\text{O}$ and percentage CaCO_3 records filtered with a band-pass gaussian filter centred on 0.0244 kyr⁻¹ (41 kyr) with a band width of 0.005 kyr⁻¹ compared with the astronomical solution for orbital obliquity^{28,29}. **c**, The palaeointensity proxy (mean NRM/IRM) placed on the $\delta^{18}\text{O}$ age model (thin line) and normalized output (thick line) of a band-pass gaussian filter centred on 0.0244 kyr⁻¹ (41 kyr) with a band width of 0.005 kyr⁻¹. **d**, $\text{IRM}_{35\text{mT}}$ placed on the $\delta^{18}\text{O}$ age model. Thick line is the normalized output of a band-pass gaussian filter centred on 0.0244 kyr⁻¹ (41 kyr) with a band width of 0.005 kyr⁻¹. **e**, The palaeointensity proxy (mean NRM/IRM) filtered with a gaussian filter centred on 0.0244 kyr⁻¹ (41 kyr) with a band width of 0.005 kyr⁻¹ compared with the astronomical solution for orbital obliquity^{28,29}.

optimal results, the remanence carrier should be fine-grained single-domain or pseudo-single-domain (PSD) magnetite, and magnetite concentration should be fairly uniform¹¹. At Site 983, the median destructive field of NRM¹², IRM acquisition¹³ (Fig. 3a) and thermal demagnetization of composite IRM¹⁴ (Fig. 3b) indicate that magnetite is the dominant magnetic mineral. Hysteresis ratios (Fig. 3c) indicate fine-grained PSD magnetite¹⁵, consistent with ARM/ χ values (where χ is volume susceptibility; Fig. 3d) which imply uniform magnetite grain size in the $\sim 1\text{--}5\text{ }\mu\text{m}$ range¹⁶. ARM and IRM values do not change by more than a factor of three or four (ref. 12), well within the criteria deemed optimal for palaeointensity determinations^{11,16}.

U-channel samples¹⁷ were collected from the archive half of the composite section at sites 983 and 984, and magnetic remanence measurements were performed on a 2G pass-through magnetometer at Gif-sur-Yvette¹⁸. U-channel measurements were made at consecutive 1-cm intervals down the section, but the width of the magnetometer response function is such that only each fourth measurement, at best, is truly independent. NRM/IRM and NRM/ARM were determined for up to five demagnetization steps in the 25–45 mT range, the ratios being computed for each peak demagnetization field. The IRM was acquired in a d.c. field of 500 mT, and the ARM in a peak alternating field of 100 mT and a 0.05 mT biasing d.c. field. The arithmetic means of NRM/IRM and of NRM/ARM were computed at 1-cm intervals down-core. Although the shape of the normalized records is very similar for

each demagnetization step, standard deviation about the mean for NRM/ARM is greater than for NRM/IRM. This observation indicates a greater discrepancy in the coercivity spectra of NRM and ARM, than NRM and IRM. NRM/IRM is, therefore, the preferred proxy for relative palaeointensity in these sediments. Mean NRM/IRM for Site 983 and Site 984 are placed on their respective timescales using oxygen-isotope stratigraphy (Figs 1c and 2c). The palaeointensity records from Site 983 and Site 984 correlate well with one another (compare Fig. 2c, d) and mismatches in NRM/IRM probably reflect uncertainties in the chronologies. For the interval 0–200 kyr ago, other high-resolution palaeointensity records—such as those from the Labrador Sea¹⁹, Mediterranean–Indian Ocean⁴ and Sulu Sea²⁰—can be matched to the Site 983 and Site 984 records; however, there are discrepancies, particularly for the past 20 kyr, which remain unresolved. Before 200 kyr ago, no records of comparable resolution are available, but correlations to the Brunhes palaeointensity record from the equatorial Pacific²¹ are possible, and the nature of the variability is comparable.

In previous time-series analysis of palaeointensity proxies, the principal shortcomings were inadequacies of age control, short record length, and low resolution of the records^{4–7}. In contrast, the Site 983 record comprises 8,220 values of mean NRM/IRM, has a relatively precise chronology and is, therefore, particularly well suited for spectral analysis. The palaeointensity data were interpolated and resampled at equal age increments maintaining about the same number of data points; the resulting interval between data

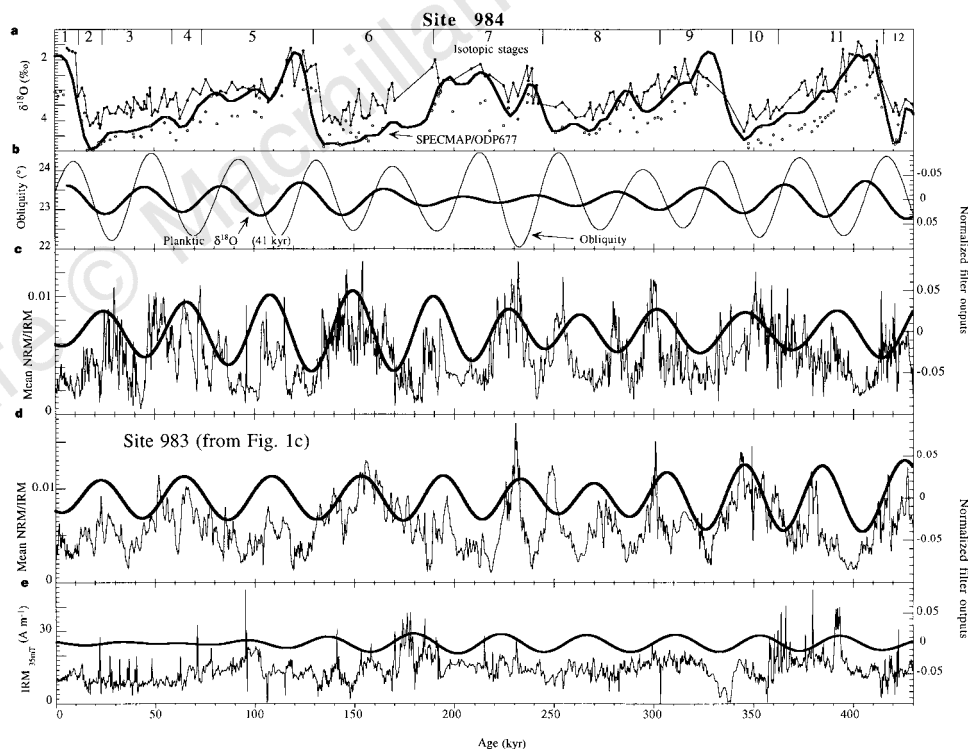


Figure 2 Site 984: relative geomagnetic palaeointensity proxy (mean NRM/IRM), the palaeointensity normalizer (IRM), and planktic and benthic oxygen-isotope data. **a**, Planktic (symbols joined by thin line) and benthic (symbols without line) oxygen-isotope ($\delta^{18}\text{O}$) records. Age models were constructed by correlation of the oxygen-isotope records to a standard reference signal (thick line) derived by combining the SPECMAP records from 0–290 kyr ago²² and 290–350 kyr (ref. 23), with the ODP Site 677 record for 350–430 kyr (ref. 26). Linear interpolation between 11 tie points (to isotopic stage 12) produced the age model. The planktic oxygen-isotope records were determined using specimens of *Neoglobobulimina pachyderma* (sinistral) selected from the 150–250 μm size fraction. The benthic oxygen-isotope record was constructed by measurement of specimens of *Cibicidoides wuellerstorfi* picked from the $>150\text{ }\mu\text{m}$ size fraction. Additional specimens of

Cibicidoides and *Uvigerina* were selected at intervals that were barren of *C. wuellerstorfi* and corrected to *C. wuellerstorfi* values. **b**, Planktic $\delta^{18}\text{O}$ record filtered with a band-pass gaussian filter centred on 0.0244 kyr^{-1} (41 kyr) with a band width of 0.005 kyr^{-1} compared with the astronomical solution for orbital obliquity^{28,29}. **c**, The palaeointensity proxy (mean NRM/IRM) placed on the $\delta^{18}\text{O}$ age model (thin line) and normalized output (thick line) of a band-pass gaussian filter centred on 0.0244 kyr^{-1} (41 kyr) with a band width of 0.005 kyr^{-1} . **d**, The palaeointensity proxy (mean NRM/IRM) for Site 983 from Fig. 1c (thin line) and normalized output (thick line) of a band-pass gaussian filter centred on 0.0244 kyr^{-1} (41 kyr) with a band width of 0.005 kyr^{-1} . **e**, $\text{IRM}_{35\text{mT}}$ placed on the $\delta^{18}\text{O}$ age model. Thick line is the normalized output of a band-pass gaussian filter centred on 0.0244 kyr^{-1} (41 kyr) with a band width of 0.005 kyr^{-1} .

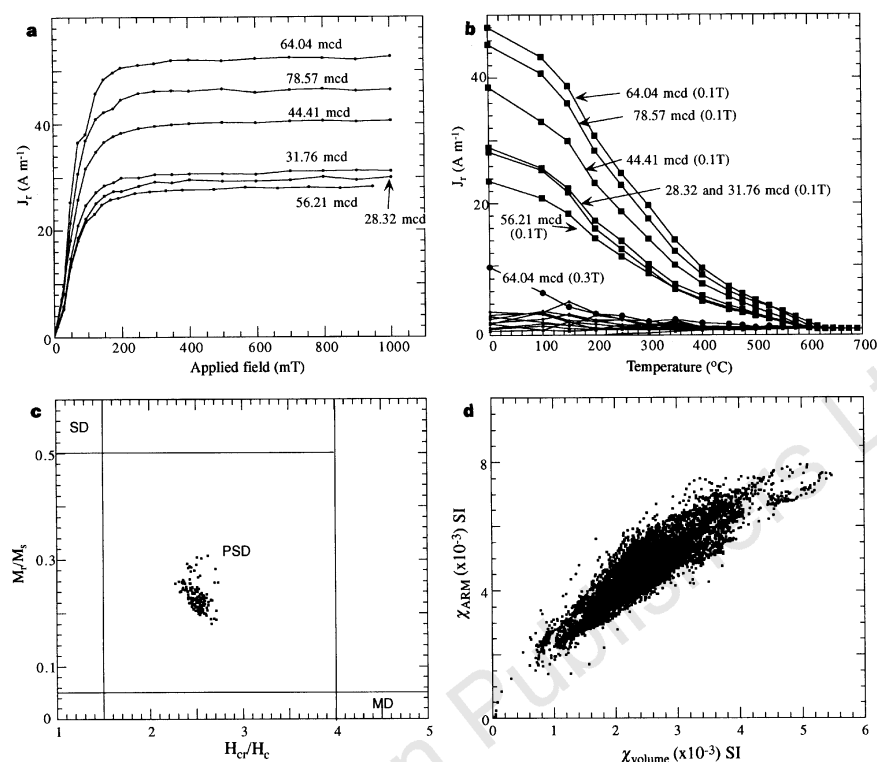


Figure 3 Site 983: rock magnetic data indicating fine-grained magnetite as the dominant magnetic mineral. **a**, Isothermal remanence (IRM) acquisition curves for samples from various depths (m.c.d., metres composite depth, J_r , remanent magnetization intensity). **b**, Thermal demagnetization of a 3-axis composite IRM¹⁴ for the same samples used in **a**. Orthogonal static fields of 1, 0.3 and 0.1 T were applied sequentially to each sample before heating. Apart from the 0.3 T IRM for sample at 64.04 m.c.d., the 0.3 T and 1 T IRMs have relatively low intensities and

blocking temperatures. **c**, Hysteresis ratios for the Brunhes chron indicating pseudo single domain (PSD) grain sizes in magnetite¹⁵. SD, single domain; MD, multidomain; H_{cr} , remanent coercivity; H_c , coercive force; M_r , saturation remanence; M_s , saturation magnetization. **d**, Anhysteretic susceptibility (χ_{ARM}) versus volume susceptibility (χ) for the Brunhes chron indicating uniformity in magnetite grain size in the $\sim 1\text{--}5\text{ }\mu\text{m}$ range¹⁶, implying suitability for relative palaeointensity determinations.

points was 88 yr and 59 yr for Site 983 and Site 984, respectively. At Site 983, both palaeointensity proxies (NRM/IRM and NRM/ARM) show power close to 0.01 kyr^{-1} (100 kyr) and 0.0244 kyr^{-1} (41 kyr) (Fig. 4a) that are significant at the 95% confidence level (Fig. 4c). Significant 41-kyr power is also apparent in records of percentage CaCO_3 at this site (Fig. 4a). The bulk magnetic properties (IRM_{35mT} and ARM_{35mT}) demagnetized at peak fields of 25 mT, and χ) all show significant power close to 0.01 kyr^{-1} (100 kyr). As the 100-kyr period is seen in all the magnetic parameters, we attribute it to lithological change associated with the 100-kyr (eccentricity) cycle that dominates glacial–interglacial climate throughout the Brunhes chron. The 0.0244 kyr^{-1} (41 kyr) power in NRM/IRM and NRM/ARM is not easily explained in terms of lithological change as it is not seen in other magnetic parameters which should monitor changes in magnetite concentration (Fig. 4b, c). At Site 984, the duration of the record is shorter than at Site 983, and the oxygen-isotope chronology is less precise. Nonetheless, 0.01 kyr^{-1} (100 kyr) and 0.0244 kyr^{-1} (41 kyr) power are also apparent in the palaeointensity proxies (Fig. 4b, d). As at Site 983, the 100-kyr power is observed in the bulk magnetic parameters (IRM_{35mT}, ARM_{35mT}, χ), and, to a lesser extent in grain-size-sensitive magnetic ratios (such as ARM/ χ and ARM/IRM); however, all these parameters are devoid of significant 41-kyr power.

The relationship between the mean NRM/IRM record and the 41-kyr period can be observed by superimposing the NRM/IRM record on the output of a band-pass gaussian filter centred on 0.0244 kyr^{-1} (41 kyr) with a band width of 0.005 kyr^{-1} (Figs 1c and 2c). The choice of band width in the $0.003\text{--}0.01\text{ kyr}^{-1}$ range does not appreciably affect the output. At both Site 983 and Site 984, lows in palaeointensity tend to coincide with highs in the angle of

obliquity (Fig. 1e), although the observed phase relationship is not constant. The same filter applied to the IRM_{35mT} record (Figs 1d and 2e) illustrates that the power in NRM/IRM is not produced by IRM. For the planktic $\delta^{18}\text{O}$ records (and the percentage CaCO_3 record for Site 983), the same filter gives outputs which indicate more or less constant phase lags with respect to orbital obliquity (Figs 1b and 2b). This is a consequence of SPECMAP age models that assume a constant lag between orbital forcing and ice-sheet response^{22–24}. The filtered NRM/IRM records are not in phase with $\delta^{18}\text{O}$ or percentage carbonate, implying that the 41-kyr power in NRM/IRM is either not climatic in origin or is a very early climatic response to orbital forcing. The former alternative is supported not only by the absence of a 41-kyr power in bulk magnetic properties, which should detect the presence of a climatically induced periodicities in magnetic mineral concentration, but also by the different phase relations of CaCO_3 (and $\delta^{18}\text{O}$) and NRM/IRM with respect to orbital obliquity (Fig. 1b, e). We conclude that orbital obliquity affects geomagnetic field intensity, presumably due to the effect of obliquity on precessional angular velocity and hence on precessional forces in the Earth's core. The lack of a constant phase relationship between 41-kyr power in palaeointensity and obliquity (Fig. 1e) may indicate variable phase shift between orbital forcing and glacial response²⁵ (whereas a constant phase shift is inherent in the SPECMAP age models). The observed variable phase shift may, however, be partially accounted for by chronological uncertainties. For example, in the 350–500 kyr ago interval at Site 983, the observed phase shift (Fig. 1e) would be reduced slightly by using SPECMAP²² rather than ODP Site 677²⁶ as the $\delta^{18}\text{O}$ reference curve. If geomagnetic palaeointensity varies in phase with obliquity, then the signal could potentially be used to improve oxygen-isotope

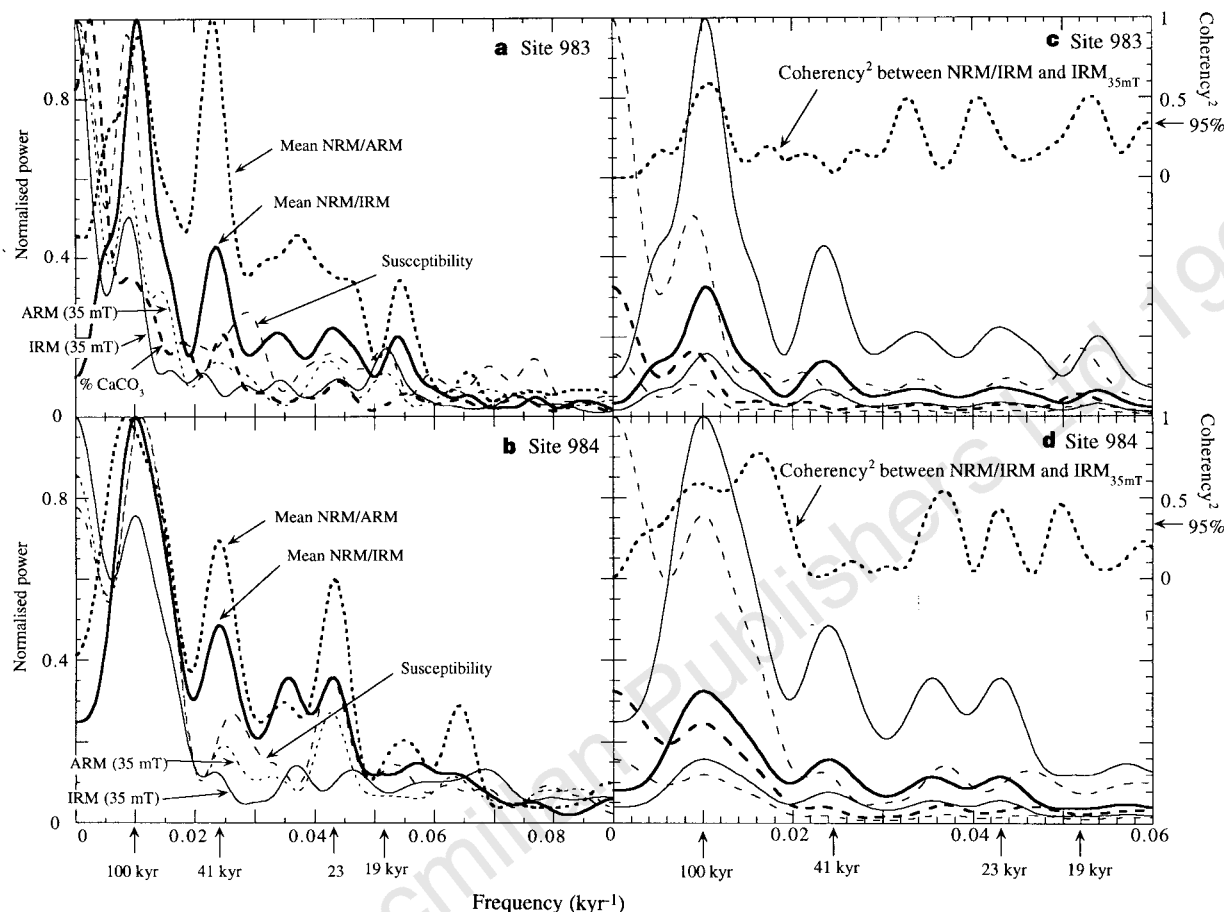


Figure 4 Power spectra and coherence analysis for various magnetic parameters, and percentage CaCO_3 using AnalySeries software³⁰ and the Blackman Tukey method^{31,32} with a Bartlett window. For Site 983 (a) and Site 984 (b), 41-kyr power is present in the palaeointensity proxies (NRM/ARM, NRM/IRM) and percentage CaCO_3 , but absent in other magnetic parameters including the

palaeointensity normalizers (ARM, IRM). For Site 983 (c) and Site 984 (d), 95% confidence limits (fine lines) about the spectra (thick lines) are shown for NRM/IRM (solid lines) and $\text{IRM}_{35\text{mT}}$ (dashed lines). Squared coherency between mean NRM/IRM and $\text{IRM}_{35\text{mT}}$ indicates insignificant coherence (at the 95% confidence level) in the 0.02–0.03 kyr^{-1} frequency interval.

chronologies, and to examine the possibility of a variable response time of global ice volume to orbital forcing.

Received 22 May 1997; accepted 1 May 1998.

- Kent, D. V. & Opdyke, N. Palaeomagnetic field intensity variations recorded in a Brunhes epoch deep-sea sediment core. *Nature* **266**, 156–159 (1977).
- Malkus, W. V. R. Precession of the Earth as a cause of geomagnetism. *Science* **160**, 259–264 (1968).
- Rochester, M. G., Jacobs, J. A., Smylie, D. E. & Chong, K. F. Can precession power the geomagnetic dynamo? *Geophys. J. R. Astron. Soc.* **43**, 661–678 (1975).
- Meynadier, L., Valet, J. P., Weeks, R., Shackleton, N. J. & Hagee, V. L. Relative geomagnetic intensity of the field during the last 140 ka. *Earth Planet. Sci. Lett.* **114**, 39–57 (1992).
- Tauxe, L. & Wu, G. Normalized remanence in sediments of the western equatorial Pacific, relative paleointensity of the geomagnetic field? *J. Geophys. Res.* **95**, 12337–12350 (1990).
- Tauxe, L. & Shackleton, N. J. Relative paleointensity records from the Ontong-Java plateau. *Geophys. J. Int.* **117**, 769–782 (1994).
- Tauxe, L. & Hartl, P. 11 million years of Oligocene geomagnetic field behaviour. *Geophys. J. Int.* **128**, 217–229 (1997).
- Shipboard Scientific Party. Site 983. *Proc. ODP Init. Rep.* **162**, 139–167 (1996).
- Shipboard Scientific Party. Site 984. *Proc. ODP Init. Rep.* **162**, 169–222 (1996).
- Hagelberg, T., Shackleton, N. J., Pisias, N. & Shipboard Scientific Party. Development of composite depth sections for Sites 844 through 854. *Proc. ODP Init. Rep.* **138**, 79–85 (1992).
- Tauxe, L. Sedimentary records of relative paleointensity of the geomagnetic field: theory and practice. *Rev. Geophys.* **31**, 319–354 (1993).
- Channell, J. E. T., Hodell, D. A. & Lehman, B. Relative geomagnetic paleointensity and $\delta^{18}\text{O}$ at ODP Site 983 (Gardar Drift, North Atlantic) since 350 ka. *Earth Planet. Sci. Lett.* **153**, 103–118 (1997).
- Dunlop, D. J. Magnetic mineralogy of unheated and heated red sediments by coercivity spectrum analysis. *Geophys. J. R. Astron. Soc.* **27**, 37–55 (1972).
- Lowrie, W. Identification of ferromagnetic minerals in a rock by coercivity and unblocking temperature properties. *Geophys. Res. Lett.* **17**, 159–162 (1990).
- Day, R., Fuller, M. & Schmidt, V. A. Hysteresis properties of titanomagnetites: grain-size and compositional dependence. *Phys. Earth Planet. Inter.* **13**, 260–267 (1977).
- King, J. W., Banerjee, S. K. & Marvin, J. A new rock-magnetic approach to selecting sediments for geomagnetic paleointensity studies: application to paleointensity for the last 4000 years. *J. Geophys. Res.* **88**, 5911–5921 (1983).
- Tauxe, L., LaBrecque, J. L., Dodson, R. & Fuller, M. U-channels—a new technique for paleomagnetic analysis of hydraulic piston cores. *Eos* **64**, 219 (1983).

- Weeks, R. *et al.* Improvements in long-core measurement techniques: applications in palaeomagnetism and palaeoceanography. *Geophys. J. Int.* **114**, 651–662 (1993).
- Stoner, J. S., Channell, J. E. T. & Hillaire-Marcel, C. Late Pleistocene relative geomagnetic paleointensity from the deep Labrador Sea: regional and global correlations. *Earth Planet. Sci. Lett.* **134**, 237–252 (1995).
- Schneider, D. A. An estimate of late Pleistocene geomagnetic intensity variation from Sulu Sea sediments. *Earth Planet. Sci. Lett.* **120**, 301–310 (1994).
- Valet, J. P. & Meynadier, L. Geomagnetic field intensity and reversals during the past four million years. *Nature* **366**, 234–238 (1993).
- Martinson, D. G. *et al.* Age dating and the orbital theory of the Ice Ages: development of a high-resolution 0 to 300,000-year chronostratigraphy. *Quat. Res.* **27**, 1–29 (1987).
- Imbrie, J. *et al.* in *Milankovitch and Climate* (eds Berger, A. L., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269–305 (NATO ASI Ser. 126, Riedel, Hingham, MA, 1984).
- Imbrie, J. & Imbrie, J. Z. Modeling the climate response to orbital variations. *Science* **207**, 943–953 (1980).
- Pisias, N. G., Mix, A. C. & Zahn, R. Nonlinear response in the global climate system: evidence from benthic oxygen isotopic record in Core RC13-110. *Paleoceanography* **5**, 147–160 (1990).
- Shackleton, N. J., Berger, A. & Peltier, W. R. An alternative astronomical calibration of the lower Pleistocene timescale based on ODP Site 677. *Trans. R. Soc. Edinb. Earth Sci.* **81**, 251–261 (1990).
- Vidal, L. *et al.* Evidence for changes in the North Atlantic deep water linked to meltwater surges during the Heinrich events. *Earth Planet. Sci. Lett.* **146**, 13–27 (1997).
- Berger, A. Long-term variations of daily insolation and Quaternary climatic change. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
- Berger, A. & Loutre, M. F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10**, 297–317 (1991).
- Paillard, D., Labeyrie, L. & Yiou, P. Macintosh program performs time-series analysis. *Eos* **77**, 379 (1996).
- Blackman, R. B. & Tukey, J. W. *The Measurement of Power Spectra from the Point of View of Communication Engineering* (Dover, New York, 1958).
- Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Applications* (Holden-Day, Oakland, CA, 1968).

Acknowledgements. We thank C. Laj, C. Kissel and A. Mazaud for logistic support and discussions; and J. S. Stoner and J. D. Ortiz for comments on the manuscript. The paleomagnetic laboratory at Gif-sur-Yvette is supported by CEA and CNRS. At the University of Florida, this project was supported by the US Science Support Program.

Correspondence and requests for materials should be addressed to J.E.T.C. (e-mail: jetc@nervm.nerdc.ufl.edu).

Energy and trace-gas fluxes across a soil pH boundary in the Arctic

D. A. Walker*, N. A. Auerbach*, J. G. Bockheim†, F. S. Chapin III‡, W. Eugster§, J. Y. King§, J. P. McFadden‡, G. J. Michaelson||, F. E. Nelson¶, W. C. Oechel#, C. L. Ping||, W. S. Reeberg§, S. Regli§, N. I. Shiklomanov¶ & G. L. Vourlitis#

* Tundra Ecosystem Analysis and Mapping Laboratory, Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado 80309-0450, USA

† Soils Department, University of Wisconsin, 1525 Observatory Drive, Madison, Wisconsin 53706, USA

‡ Department of Integrative Biology, University of California Berkeley, Berkeley, California 94720, USA

§ Department of Earth System Science, University of California Irvine, 205 Physical Sciences, Irvine, California 92697-3100, USA

|| Agricultural and Forestry Exploratory Station, University of Alaska Fairbanks, 533 E. Fireweed, Palmer, Arkansas 99645, USA

¶ Department of Geography, University of Delaware, Newark, Delaware 19716, USA

Global Change Research Group, Department of Biology, San Diego State University, San Diego, California 92182, USA

Studies and models of trace-gas flux in the Arctic consider temperature and moisture to be the dominant controls over land–atmosphere exchange^{1,2}, with little attention having been paid to the effects of different substrates. Likewise, current Arctic vegetation maps for models of vegetation change recognize one or two tundra types^{3,4} and do not portray the extensive regions with different soils within the Arctic. Here we show that rapid changes to ecosystem processes (such as photosynthesis and respiration) that are related to changes in climate and land usage will be superimposed upon and modulated by differences in substrate pH. A sharp soil pH boundary along the northern front of the Arctic Foothills in Alaska separates non-acidic (pH > 6.5) ecosystems to the north from predominantly acidic (pH < 5.5) ecosystems to the south. Moist non-acidic tundra has greater heat flux, deeper summer thaw (active layer), is less of a carbon sink, and is a smaller source of methane than moist acidic tundra.

In 1995 and 1996, we studied the ecosystem properties on either side of a prominent pH boundary within the Kuparuk River basin (KRB) in Alaska, the primary study area of the Arctic System Science Land–Atmosphere–Ice Interactions (ARCSS–LAII) Flux Study⁵ (Fig. 1). We characterized moist non-acidic tundra (MNT) and moist acidic tundra (MAT) ecosystems at two intensive study sites about 7 km apart on either side of the boundary (Fig. 1b, sites 3 and 4). We also collected soil and vegetation data from numerous other MNT and MAT sites within the KRB during an accuracy assessment of the landcover map in Fig. 1b (ref. 6). This adds to earlier information from Toolik Lake, Happy Valley and Prudhoe Bay, Alaska^{7–11}.

The vegetation and soil properties on either side of the boundary are similar to those described for MNT and MAT in other studies^{8,12}. Site 3 has MNT with 36% cover of non-sorted circles¹³. The non-sorted circles are partly vegetated patches of highly frost-active soils that are about 1–2 m in diameter and spaced at intervals of 2–3 m; bare soil covers about 4% of site 3. The vegetation community between the circles is *Dryado integrifoliae-Caricetum bielowiei*⁸, which is dominated by non-tussock sedges (*Carex bigelowii*, *C. membranacea* and *Eriophorum triste*), prostrate shrubs (*Dryas integrifolia*, *Salix arctica*, *S. reticulata* and *Arctous rubra*) and minerotrophic mosses (*Tomentypnum nitens*, *Hylocomium splendens* and *Ditrichum flexicaule*). Soils of MNT have a broken organic layer over a dark-coloured A horizon (a mineral horizon containing

organic-matter accumulation) with high base saturation, over a gleyed C horizon (a subsoil mineral horizon relatively unaffected by soil-formation processes except for the presence of grey colours resulting from poor drainage and reduction of iron)^{11,14}. All soil horizons have consistently high pH (>6.5) and are highly frost stirred (cryoturbated).

Site 4 is covered by tussock tundra (*Sphagno-Eriophoretum*⁸) with few (<1% cover) non-sorted circles. This vegetation type is dominated by dwarf shrubs (*Betula nana*, *Ledum palustre* ssp. *decumbens*, *Salix planifolia pulchra*), tussock sedges (*Eriophorum vaginatum*) and acidophilous mosses (*Sphagnum* spp., *Aulacomnium* spp., *Polytrichum* spp. and *Dicranum* spp.). Soils of MAT have a thick continuous organic horizon over gleyed subsoil material and contain cryoturbated organic material in the lower part. Both sites 3 and 4 are on silty loess deposits¹¹. Soil pH of MAT sites tends to increase with depth from about 4.0 at the surface to 6.5 in the frozen C horizons.

The pH boundary extends at least 300 km to the east and west of the study area^{15,16}. Loess blankets much of the Arctic Coastal Plain and Arctic Foothills, and both MAT and MNT occur on these extensive deposits, so it is difficult to explain the sharp vegetation boundary solely by differences in surface deposits¹⁷. The boundary may be partly due to a stronger winter Arctic climate north of the topographic barrier of the Arctic Foothills¹⁸. A colder, windier climate with shallower snowpack would promote the formation of non-sorted circles¹² and cause the continual stirring of non-acidic subsoils to the surface^{11,14}. The abundance of non-sorted circles and relatively low shrub biomass (85 versus 202 g m⁻²) north of the boundary results in the greyer tones on the false-colour infrared image (Fig. 1a). Lower shrub biomass, lower leaf-area index (LAI) and lower normalized difference vegetation index (NDVI) of MNT at site 3 is consistent with previous studies^{9,19} (Table 1).

South of the boundary, MNT is found only in relatively small areas on limestone bedrock and in naturally disturbed systems, such as river floodplains, snowbeds, windy hill crests and recently glaciated areas. In most of the Arctic Foothills, vegetation succession and peat formation (paludification) during the Holocene have converted formerly dry vegetation on mineral-rich loess and till deposits to MAT. Paludification is enhanced toward the south as a result of increased temperature and precipitation. Mosses, particularly *Sphagnum*, are important to this conversion. It is abundant in MAT but not MNT, and has numerous unique properties that strongly promote waterlogging and cold acidic soils^{20–23}.

The vegetation and soil differences between MAT and MNT have important consequences for land–atmosphere exchanges. Site 3 (MNT) had 28% more soil heat flux during 10 days of observation and 54% deeper end-of-summer thaw than site 4 (MAT). Summer thaws of MNT are consistently deeper than those of MAT throughout the KRB, despite MNT being dominant in the northern, colder portion of the study area²⁴, because the MNT has shorter, more open plant canopies (less shading by vascular-plant leaf area), less continuous moss cover and thinner organic horizons (Table 1). In a related study, evapotranspiration, soil heat flux and sensible heat flux (heat exchange between the atmosphere and the land surface) showed a similar relationship with net radiation at two acidic tundra sites (sites 4 and 6; Fig. 1b) despite latitudinal and elevation differences in climate, indicating that the energy budgets are more strongly correlated with vegetation type than with climate²⁵.

Site 4 also had about twice the gross photosynthesis and three times the respiration of the MNT site, as well as a greater net carbon gain, during the same 10-day measurement period (Table 1), despite the close proximity of the two sites and nearly identical temperature, net radiation and relative humidity²⁵. These results are consistent with CO₂-flux data from two other sites (11 and 21; Fig. 1b) during the same period in 1995. Site 11 (MAT) had similar summer climate and CO₂ flux to that at site 4, whereas site 21 (MNT) had a lower flux than site 3, probably owing to the colder early summer climate near the coast²⁶. Integrated fluxes from sites 11 and 21

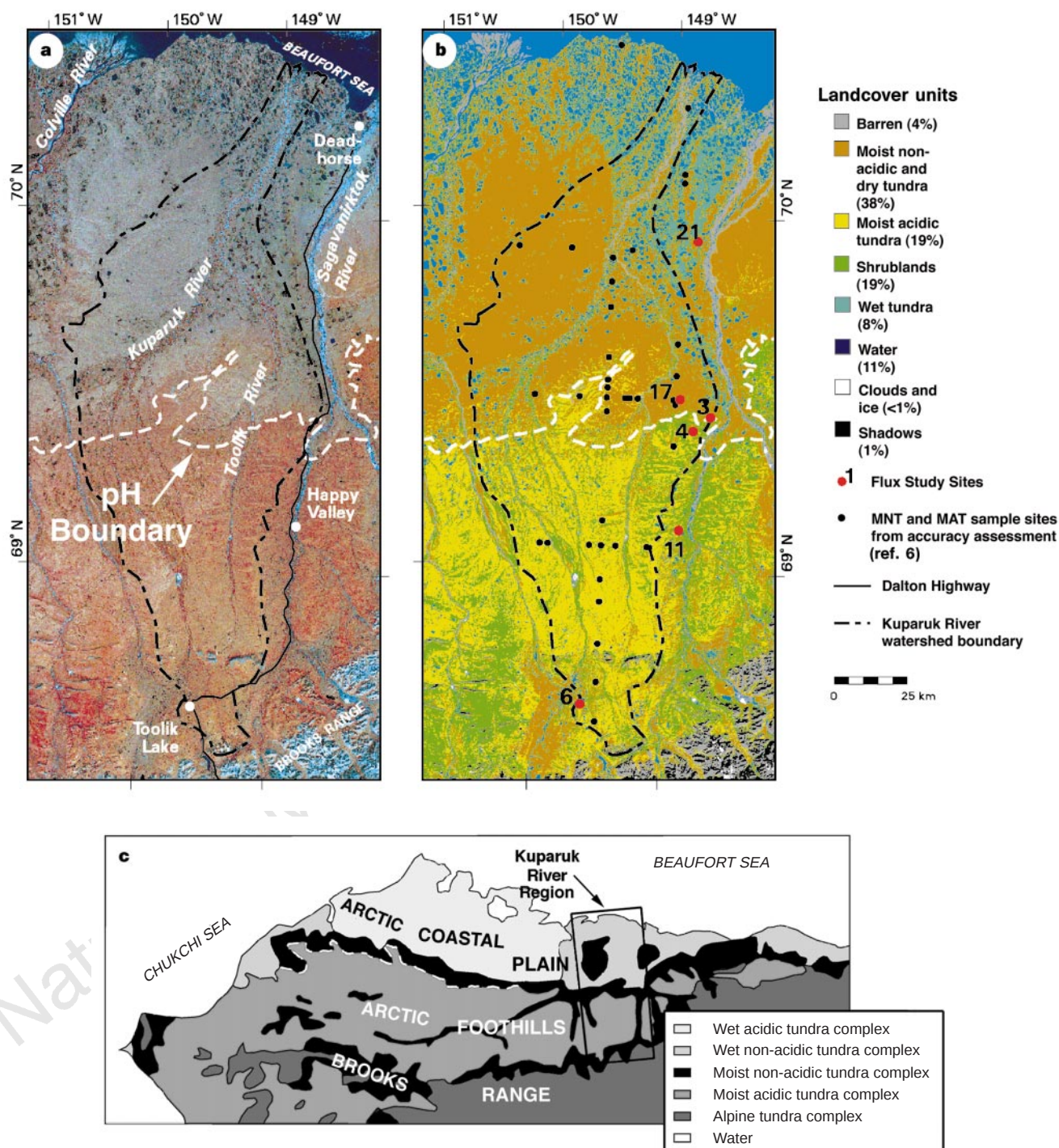


Figure 1 Maps of the study area. **a**, Landsat MSS false-colour infrared mosaic of the Kuparuk River Basin (dashed line), northern Alaska. A broad band of foothills occurs between the mountains and the Arctic Coastal Plain with many lakes. The pH boundary separating the redder tones to the south from the greyer tones to the north corresponds to soils that are acidic and non-acidic, respectively. Landsat data are courtesy of the US Geological Survey Alaska Data Center. **b**, Land-cover map. Red circles indicate intensive study sites. Black dots are other sites visited

throughout the summer of 1995 showed that the MAT site was a much greater carbon sink than the MNT site (55.2 versus 27.6 g C m⁻² per season). In 1996, an even larger difference was observed between site 11 and site 17, an MNT site close to site 3 that has a summer climate very similar to sites 3 and 11. Site 11 (MAT) gained 52.5 g C m⁻² per season compared with 3.3 g C m⁻² per season at site 17 (MNT) (Table 1). Taken together, our data demonstrate a consistent spatial and temporal pattern of a much larger carbon sink

during the accuracy assessment of the map⁶. **c**, The map shows a generalized distribution of acidic and non-acidic vegetation types in northern Alaska, prepared from an integration of information from several sources including AVHRR (Advanced Very-High Resolution) satellite images, soil maps, vegetation maps and surface-geology maps. The location of the pH boundary west of the Colville River (white dashed line) is less distinct and unstudied.

in MAT than in MNT. Methane flux showed a pattern opposite to that of CO₂, with the wetter, more anaerobic soils of MAT effluxing over six times the methane of MNT (Table 1).

Greater carbon accumulation in the vegetation has also led to twice as much organic carbon in both the active layer and the permafrost in the soils at site 4 than site 3 (Table 1). The basal ¹⁴C date from the frozen C horizon at site 4 was 8,500 years BP compared to 12,500 years BP in the same horizon at site 3, demonstrating the

Table 1 Comparison of ecosystem properties of MNT and MAT at sites in the Kuparuk River Basin

Ecosystem property	Sites 3 and 4			Other sites			Reference
	MNT	MAT	Significance	MNT	MAT	Significance	
Soil							
pH of top mineral horizon	7.6 [1]	5.5 [1]	n.a.	7.0 ± 0.16 [20] 6.3 ± 0.1 [14]	5.3 ± 0.13 [10] 4.6 ± 0.1 [33]	***	Ref. 14 Ref. 8†
O-horizon thickness (cm)	9 ± 1 [71]	15 ± 1 [71]	**	11 ± 1.9 [21]	21 ± 1.8 [15]	***	This study‡ This study§
Soil moisture of top mineral horizon (cm³ cm⁻³, Jul 95)	0.37 [1]	0.40 [1]	n.a.				
Bare soil (% cover)	4.4 ± 1.6 [6]	0.2 ± 0.0 [6]	***	8 ± 1 [140]	1 ± 0.2 [121]	***	This study§
Vegetation							
Height of plant canopy (cm)	3.9 ± 0.3 [340]	6.5 ± 0.4 [340]	***				This study
Leaf area index	0.50 ± 0.03 [66]	0.84 ± 0.05 [66]	***	0.57 ± 0.06 [7]	0.81 ± 0.08 [11]	**	Ref. 19
NDVI (MSS)	0.23 [1]	0.32 [1]	n.a.	0.28 ± 0.00 [4 × 10⁶]	0.41 ± 0.00 [2 × 10⁶]	n.a.	This study¹¹
NDVI (Hand-held)				0.62 ± 0.02 [7]	0.71 ± 0.01 [11]	**	Ref. 19
Moss cover (%)	65 ± 4 [12]	79 ± 4 [12]	**				This study
Above ground biomass (g m⁻²)							
Shrubs	85 ± 18 [10]	202 ± 22 [10]	***	127 ± 19 [7]	270 ± 19 [11]	***	Ref. 19
Graminoids	124 ± 12 [10]	112 ± 15 [10]	n.s.	118 ± 22 [7]	118 ± 24 [11]	n.s.	Ref. 19
Forbs	40 ± 22 [10]	10 ± 2 [10]	***	12 ± 4 [7]	12 ± 2 [11]	n.s.	Ref. 19
Mosses, lichens, litter	504	460	n.a.	221 ± 85 [7]	207 ± 33 [11]	n.s.	Ref. 19
Total	753 ± 60 [10]	784 ± 139 [10]	n.s.	447 ± 23 [7]	607 ± 27 [11]	***	Ref. 19
Energy and trace-gas flux							
Soil heat flux (19–30 Jun 1995, MJ m⁻² d⁻¹)	1.39 ± 0.21 [331]	1.09 ± 0.16 [275]	***				This study
Thaw depth (cm)	57 ± 1 [71]	37 ± 1 [71]	***	52 ± 2 [20] 57 ± 5 [14]	39 ± 2 [14] 36 ± 3 [33]	***	This study§ Ref. 8
Evapotranspiration (19–30 Jun 1995, mm d⁻¹)	1.16 ± 0.17 [331]	1.06 ± 0.16 [275]	n.a.				This study
10-d gross primary production (19–30 Jun 1995 g CO₂-C m⁻² d⁻¹)	0.94 ± 0.14 [331]	1.82 ± 0.27 [275]	n.a.				This study
10-d net CO₂ uptake (g CO₂-C m⁻² d⁻¹)	0.67 ± 0.10 [331]	0.95 ± 0.27 [275]	n.a.	0.27 ± 0.41 [12]	1.02 ± 0.33 [12]	n.a.	This study¶
10-d respiration loss (g CO₂-C m⁻² d⁻¹)	0.27 ± 0.04 [331]	0.87 ± 0.13 [275]	n.a.				This study¶
1995 net CO₂ uptake (g CO₂-C m⁻² per season)				27.6 [77]	55.2 [90]	n.a.	This study#
1996 net CO₂ uptake (g CO₂-C m⁻² per season)				3.3 [31]	52.5 [73]	n.a.	This study*
Methane emission (mg CH₄ cm⁻² yr⁻¹)				69 ± 33 [12]	449 ± 301 [15]	*	This study††
Soil organic carbon (kg C m⁻³)	40 [1]	88 [1]	n.a.	56 ± 5 [5] 55 ± 5 [16]	44 ± 11 [6] 49 ± 4 [7]	n.s. n.s.	Ref. 11 Ref. 10

Standard error of the mean and number of samples [in brackets] are given for most variables. Probability of significance in all cases was based on two-sample *t*-test. Significance levels:

* $P \leq 0.1$; ** $P \leq 0.05$; *** $P \leq 0.01$; n.s., non-significant; n.a., non-applicable.

† Data are from 47 permanent plots in the Toolik Lake region.

‡ Measurements at 36 random points within the Kuparuk River basin during accuracy assessment of the land-cover map.

§ Estimates obtained from aerial surveys at 361 sites within the Kuparuk River basin during accuracy assessment of the land-cover map.

|| Mean MSS NDVI values for the land-cover map.

¶ Sites 11 (MAT) and 24 (MNT).

Sites 11 (MAT) and 24 (MNT). Time intervals for measurements: site 11, 1 Jun–31 Aug 1995; site 24, 16 Jun–31 Aug 1995.

* Sites 11 (MAT) and 17 (MNT). Time intervals for measurements: site 11, 6 Jun–31 Aug 1996; site 17, 15 Jun–19 Aug 1996.

†† Methane measurements at 37 plots at Toolik Lake region, Happy Valley and Deadhorse.

faster accumulation rate at the acidic site. Other soil data from MAT and MNT sites throughout the basin do not show a similar consistent trend of more carbon in the MAT soils (Table 1), presumably because there is a wide diversity of genetic environments, including MNT fen and fluvial sites, and MAT sites on a variety of surface ages. Other studies, however, show that MNT soils have consistently lower C:N ratios (15 versus 20)¹¹, greater microbial activity and more highly decomposed organic fraction²⁷. The relative winter CO₂ flux rates of MNT and MAT remain unresolved²⁸.

Extrapolations of trace-gas fluxes and soil carbon based solely on numbers from the more extensively studied MAT, as has been done in all previous high-latitude extrapolations, results in large errors. For example, in the map area of Fig. 1, this would overestimate gross photosynthesis by at least 35%, respiration by 140%, net CO₂ uptake by at least 15%, and methane flux by 140%. Similar, but more diffuse, pH boundaries separate worldwide zonal tundra types. MNT corresponds to the sedge-dominated 'typical tundra' of Russian authors, whereas MAT corresponds to shrubby 'southern tundra'¹². Over century to millennium time scales, we expect that zonal soil pH boundaries will shift northwards in response to climate warming, deeper winter snowpacks and reduction of loess sources. Regions with declining soil pH will show a decrease in soil heat flux and large increases in methane flux and carbon storage in the plant canopy. □

Methods

Site selection. In the summers of 1995 and 1996, several sites were monitored

to characterize the energy and trace-gas fluxes of arctic tundra⁵. Sites 3, 17 and 24 have MNT and sites 4, 6 and 11 have MAT vegetation. To compare MNT and MAT fluxes under nearly identical summer climates, we chose two sites about 7 km apart (sites 3 and 4; Fig. 1b) on opposite sides of the pH boundary on hilltops (224 and 332 m, respectively) with similar topography. The sites were accessible by helicopter from Happy Valley and were as similar and homogeneous as possible, and were far enough from the Dalton Highway to eliminate the effects of road dust.

Soils. Percentages of soil types and O-horizon thickness at sites 2 and 4 were determined from 71 random points at each site. Soil classification is according to the Gelisol Order in US soil taxonomy²⁹. The pH values at all sites are from surface samples collected at 39 random MNT and 24 MAT points within the KRB basin. The mean O-horizon thickness for the same sites was determined from 10 samples at each site. Bare-soil values for the KRB are visual estimates from 261 sites visited during accuracy assessment of the land-cover map⁶. Soil organic carbon was analysed on acid-treated samples using a Leco CHN-1000 analyser¹¹.

Vegetation. The land-cover map of the KRB was derived from a mosaic of Landsat Multispectral Scanner (MSS) images provided by the USGS EROS Alaska Data Center⁶. The height of the plant canopy was determined from 340 random points each at sites 3 and 4. Percentage cover of moss and bare soil was determined from four 50-m and two 70-m line transects at each site. Vascular plant leaf-area index was measured with a LI-COR PCA 2000 plant canopy analyser¹⁹ at 66 points at each site. Biomass is the mean dry mass of 10 random 20 × 50-cm clip harvest plots within sites 3 and 4. The normalized difference vegetation index (NDVI) at sites 3 and 4 was determined from the pixel of the Landsat MSS image centred on the sample sites. Mean NDVI for MNT and MAT

within the KRB was calculated from the total set of pixels in each class in Fig. 1.

Flux measurements. Short-term flux measurements were made simultaneously at sites 3 and 4 from 19 Jun to 29 Jun 1995, using four heat-flux plates and four temperature probes. Evapotranspiration and CO₂ flux were measured using the eddy-covariance method with an Applied Technologies sonic anemometer and LI-COR 6262 infrared gas analyser mounted on 2-m towers²⁵. The mean and standard error for energy flux, gross primary production and evapotranspiration at sites 3 and 4 were calculated on the basis of 30-min averages. CO₂ fluxes at sites 11, 17 and 21 were determined using eddy-covariance methods and 2.5-m towers²⁶. Mean values and standard errors at these sites were calculated using the daily mean CO₂ fluxes. The daily methane fluxes were integrated over the thaw period to obtain annual emission. Winter methane fluxes were assumed to be zero. CH₄ flux was measured during the thaw season, Jun–Aug, at 27 MNT and MAT sites along the Dalton Highway in 1996 using a static chamber method³⁰. Air samples were taken over periods of 30–45 min and were analysed on a gas chromatograph equipped with a flame ionization detector.

Received 27 February; accepted 16 April 1998.

- Rastetter, E. B., McKane, R. B., Shaver, G. R., Nadelhoffer, K. J. & Giblin, A. in *Global Change and Arctic Terrestrial Ecosystems* (eds Oechel, W. C. et al.) 437–451 (Springer, New York, 1997).
- Vourlitis, G. L. & Oechel, W. C. in *Global Change and Arctic Terrestrial Ecosystems* (eds Oechel, W. C. et al.) 266–289 (Springer, New York, 1997).
- Lynch, A., Chapman, W., Walsh, J. & Weller, G. Development of a regional climate model of the Western Arctic. *J. Clim.* **8**, 1556–1570 (1995).
- Prentice, I. C. et al. A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* **19**, 117–134 (1992).
- Weller, G. et al. The Arctic Flux Study: a regional view of trace gas release. *J. Biogeogr.* **22**, 365–374 (1995).
- Muller, S. V. et al. Accuracy assessment of a land-cover map of the Kuparuk River basin, Alaska: Considerations for remote regions. *Photogram. Eng. Remote Sens.* (in the press).
- Walker, D. A. & Everett, K. R. Loess ecosystems of northern Alaska: regional gradient and toposequence at Prudhoe Bay. *Ecol. Monogr.* **61**, 437–464 (1991).
- Walker, M. D., Walker, D. A. & Auerbach, N. A. Plant communities of a tussock tundra landscape in the Brooks Range Foothills, Alaska. *J. Veg. Sci.* **5**, 843–866 (1994).
- Walker, D. A., Auerbach, N. A. & Shippert, M. M. NDVI, biomass, and landscape evolution of glaciated terrain in northern Alaska. *Polar Rec.* **31**, 169–178 (1995).
- Bockheim, J. G., Walker, D. A. & Everett, L. R. in *Advances of Soil Science. Proc. Int. Symp. Carbon Sequestration in Soil* (eds Lal, R., Kimble, J. M., Follett, R. F. & Stewart, B. A.) 143–155 (CRC Press, Boca Raton, FL, 1996).
- Michaelson, G. J., Ping, C. L. & Kimble, J. M. Carbon storage and distribution in tundra soils of Arctic Alaska. *U.S.A. Arct. Alp. Res.* **28**, 414–424 (1996).
- Matveyeva, N. V. *Zonation in Plant Cover of the Arctic* (Russian Academy of Sciences, Proceedings of the Komarov Botanical Institute, No. 21, 1998).
- Washburn, A. L. *Geocryology: A Survey of Periglacial Processes and Environments* (Halsted/Wiley, New York, 1980).
- Bockheim, J. G., Walker, D. A., Everett, L. R., Nelson, F. E. & Shiklomanov, N. I. Soils and cryoturbation in moist nonacidic and acidic tundra in the Kuparuk river basin, arctic Alaska. *U.S.A. Arct. Alp. Res.* (in the press).
- Ebersole, J. J. *Vegetation Disturbance and Recovery at the Oumalik Oil Well, Arctic Coastal Plain, Alaska*. Thesis, Univ. Colorado, Boulder (1985).
- Walker, D. A. et al. *Landsat-assisted Environmental Mapping in the Arctic National Wildlife Refuge, Alaska* (U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, NH, CRREL Report 82-37, 1982).
- Carter, L. D. A Pleistocene sand sea on the Alaskan Arctic Coastal Plain. *Science* **211**, 381–383 (1981).
- Zhang, T., Osterkamp, T. E. & Stamnes, K. Some characteristics of the climate in Northern Alaska. *U.S.A. Arct. Alp. Res.* **28**, 509–518 (1996).
- Shippert, M., Walker, D. A., Auerbach, N. A. & Lewis, B. E. Biomass and leaf-area index maps derived from SPOT images for Toolik Lake and Imnavait Creek areas, Alaska. *Polar Rec.* **31**, 147–154 (1995).
- Benninghoff, W. S. Interaction of vegetation and soil frost phenomena. *Arctic* **5**, 34–44 (1952).
- Heinselman, M. L. Landscape evolution, peatland types, and the environment in the Lake Agassiz Peatlands Natural Area, Minnesota. *Ecol. Monogr.* **40**, 235–261 (1970).
- Sjörs, H. Divergent succession in mires, a comparative study. *Aquilo Ser. Bot.* **28**, 67–77 (1990).
- Klinger, L. F. Coupling of soils and vegetation in peatland succession. *Arct. Alp. Res.* **28**, 380–387 (1996).
- Nelson, F. E. et al. Estimating active-layer thickness over large regions: Kuparuk River Basin, Alaska. *Arct. Alp. Res.* **29**, 367–378 (1997).
- Eugster, W., McFadden, J. P. & Chapin, F. S. I. A comparative approach to regional variation in surface fluxes using mobile eddy correlation towers. *Boundary-Layer Meteorol.* **85**, 293–307 (1997).
- Vourlitis, G. L. & Oechel, W. C. Landscape-scale CO₂, H₂O vapour and energy flux of moist-wet coastal tundra ecosystems over two growing seasons. *J. Ecol.* **85**, 575–590 (1997).
- Letteme, D., Ping, C. L. & Michaelson, G. J. Effect of temperature, soil, and permafrost on microbial respiration of CO₂ from acidic and nonacidic tundra. *LAI Science Workshop* (LAI Science Management Office, University of Alaska Fairbanks, 1997).
- Oechel, W. C., Vourlitis, G. & Hastings, S. J. Cold-season CO₂ emission from arctic soils. *Global Biogeochem. Cycles* **11**, 163–172 (1997).
- Bockheim, J. G. & Tarnocai, C. in *Handbook of Soil Science* (eds Sumner, M. E. et al.) (CRC Press, Boca Raton, FL, in the press).
- Whalen, S. C. & Reeburgh, W. S. A methane flux transect along the trans-Alaska pipeline haul road. *Tellus* **42B**, 237–249 (1990).

Acknowledgements. The following US National Science Foundation projects contributed to this work: Arctic Transitions in the Land-Atmosphere System, the Flux Study, the Arctic Long-Term Ecology Research project, and the International Tundra Experiment. The US Department of Energy's Response, Resistance, Resilience, and Recovery of Arctic Ecosystems to Disturbance project also made significant contributions to the data used in the project.

Correspondence and requests for materials should be addressed to D.A.W. (e-mail: swalker@taimyr.colorado.edu).

Complementarity and the use of indicator groups for reserve selection in Uganda

Peter C. Howard*†, Paolo Viskanic*†, Tim R. B. Davenport*†, Fred W. Kigenyi*, Michael Baltzer*†, Chris J. Dickinson*†, Jeremiah S. Lwanga*†, Roger A. Matthews*† & Andrew Balmford‡

* Forest Department, PO Box 1752, Kampala, Uganda

‡ Department of Animal and Plant Sciences, The University of Sheffield, Sheffield S10 2TN, UK

A major obstacle to conserving tropical biodiversity is the lack of information as to where efforts should be concentrated. One potential solution is to focus on readily assessed indicator groups, whose distribution predicts the overall importance of the biodiversity of candidate areas^{1,2}. Here we test this idea, using the most extensive data set on patterns of diversity assembled so far for any part of the tropics. As in studies of temperate regions^{2–8}, we found little spatial congruence in the species richness of woody plants, large moths, butterflies, birds and small mammals across 50 Ugandan forests. Despite this lack of congruence, sets of priority forests selected using data on single taxa only often captured species richness in other groups with the same efficiency as using information on all taxa at once. This is because efficient conservation networks incorporate not only species-rich sites, but also those whose biotas best complement those of other areas^{9–11}. In Uganda, different taxa exhibit similar biogeography, so priority forests for one taxon collectively represent the important forest types for other taxa as well. Our results highlight the need, when evaluating potential indicators for reserve selection, to consider cross-taxon congruence in complementarity as well as species richness.

By containing elements of both East African savannas and Central African rain forests, Uganda boasts more species for its size than almost any other country in Africa¹². Much of this diversity is restricted to 15,000 km² of forest reserves (which also contain non-forest habitats) under the jurisdiction of the Uganda Forest Department¹³. The aim of a five-year inventory of the woody plants, large moths (saturnids and sphingids), butterflies, birds, and small mammals (rodents and insectivores) of all of the principal forest reserves was to provide information to the government regarding a plan to protect ~3,000 km² (20%) of the remaining forest estate as a strict nature reserve^{14,15}. Forests were surveyed in proportion to their area (see Methods). In total, nearly 100 man-years of survey effort yielded records of 2,452 species.

Constraints on funding and expertise mean that surveys of this magnitude will rarely be undertaken elsewhere in the tropics. However, the size and taxonomic breadth of the Uganda data set mean that it provides an exceptional opportunity to test ways in which future priority-setting exercises could be conducted more quickly and at lower cost. Here we focus on one widely proposed short cut to establishing priorities for biodiversity conservation, and determine whether survey data on just one or two putative indicator groups can identify robust reserve networks capable of conserving biodiversity as a whole^{1,2}.

† Present addresses: Wildlife Department, PO Box M239, Accra, Ghana (P.C.H.); Food and Agriculture Organisation, PO Box 521, Kampala, Uganda (P.V.); World Wide Fund for Nature-Cameroon, PO Box 6776, Yaounde, Cameroon (T.R.B.D.); Groundtruth International Ltd, 4 Duck Lane, Oundle, Northants PE8 4DY, UK (M.B.); Wildlife Conservation Division, Royal Forest Department, Phaholoythin Road, Chantachak, Bangkok 10900, Thailand (C.J.D.); Makerere University Biological Field Station, PO Box 409, Fort Portal, Uganda (J.S.L.); and Countryside Council for Wales, Park Street, Newtown, Powys SY16 1RD, UK (R.A.M.).

Table 1 Cross-taxon congruence

	Large moths	Butterflies	Birds	Small mammals
Total recorded species richness				
Woody plants	0.68 $P < 0.001$	0.72 $P < 0.001$	0.69 $P < 0.001$	0.53 $P < 0.001$
Large moths		0.76 $P < 0.001$	0.75 $P < 0.001$	0.53 $P < 0.001$
Butterflies			0.68 $P < 0.001$	0.49 $P < 0.001$
Birds				0.66 $P < 0.001$
Residual species richness				
Woody plants	0.13 NS	0.14 NS	0.07 NS	0.12 NS
Large moths		0.55 $P < 0.001$	0.32 $P < 0.05$	0.22 NS
Butterflies			0.34 $P < 0.05$	0.04 NS
Birds				0.25 $P < 0.10$
Complementarity scores				
Woody plants	0.77 $P < 0.001$	0.74 $P < 0.001$	0.85 $P < 0.001$	0.51 $P < 0.001$
Large moths		0.68 $P < 0.001$	0.71 $P < 0.001$	0.52 $P < 0.001$
Butterflies			0.75 $P < 0.001$	0.58 $P < 0.001$
Birds				0.52 $P < 0.001$

Cross-taxon congruence in total recorded species richness ($N = 50$ forests), residual species richness after controlling for sampling effort ($N = 50$ forests), and complementarity scores ($N = 49$ pairs of forests). Values are Pearson correlation coefficients. NS, not significant.

One simple route to identifying possible indicator taxa is to quantify how far spatial patterns of species richness coincide across different groups^{2,3,5,6,16–19}. High congruence would be extremely interesting from an evolutionary perspective, and encouraging in terms of rapid assessment of biodiversity, but so far studies of temperate areas have revealed rather low congruence in species richness^{3,4,7,8}. Because of increased endemism, there might not be such low congruence in the tropics, where high biological diversity coupled with very limited resources for its assessment mean that the benefits of using indicators would be greatest. However, no equivalent analyses at scales appropriate for reserve planning have been conducted yet anywhere in the tropics (but see ref. 18).

In practice, our data from Uganda at first suggest high congruence: there is a good match across taxa in the total numbers of species recorded from each forest (Table 1). But this apparent congruence is driven largely by differences in forest area and hence sampling effort, rather than congruence in species richness per unit area. Because of survey design, sampling effort per forest was closely correlated across taxa ($P < 0.001$ for all ten Pearson correlations), with larger, more extensively sampled forests generally appearing species-rich and smaller forests appearing species-poor as a result. Controlling for differences in sampling effort by first taking residuals from regressions of total species richness on the number of days spent sampling a taxon, and then comparing these residual scores across pairs of taxa, shows that there is low underlying congruence in species richness (only 3 out of 10 cross-taxon correlations are significant at $P < 0.05$; Table 1). Controlling for variation in forest size and sampling effort in other ways (such as taking residuals about forest area, or calculating relative species richness after conducting rarefaction²⁰) confirmed that congruence in underlying species richness is very low. Our results thus mirror earlier findings, and, taken in isolation, indicate that concentrating on one or a few groups may be of limited utility in setting site-based priorities for conserving biodiversity as a whole.

However, looking at congruence in species richness is not a sufficient test of the ability of one taxon to indicate the overall conservation value of different sites^{2,5,21–23}. A more appropriate test is to see how well sets of priority sites chosen using information from just that group capture diversity in other taxa as well. We used

a simple but powerful complementarity-based algorithm to identify networks of forests containing near-maximum numbers of species in a fixed area (nominally 20% of the total forest estate; see Methods). We ran this algorithm using data either from single groups or from all taxa at once. For each of the resulting site-selection sequences, we examined the manner in which the cumulative number of species in all survey groups increased as a function of the combined area of the sites selected.

The top 20% of forest area picked using only the data on butterflies or birds contained as many or more species overall as the equivalent set of sites identified using data on all groups together (whether measured as the mean percentage representation of all five groups (Fig. 1), or as a cumulative representation of all species (not shown)). The data on moths also performed well, although using data on the poorest-performing group (which, at the 20% threshold, is the plants) resulted in the loss of ~10% of species, compared to using data on all taxa.

One explanation for these results could be that nearly any selection sequence simply reflects forest area (with the biggest forests being chosen first), indicating that priority sets of forests could perhaps be identified using information on forest size alone. But when we tested this by selecting sites simply in order of decreasing area, the resulting networks were very inefficient: choosing the top 20% of the forest estate this way resulted in the loss of 400 or more species, compared with using data on birds or butterflies. One last, important check is to examine the efficiency of picking sites entirely at random. On average (across 100 runs each for networks of 1, 2, 3 ... 50 sites), randomly selected networks contained two to three times more sites, of smaller average size, than systematically picked networks of the same total area. Despite this, random selection captured fewer species per unit area than using data on birds, butterflies or all groups of taxa (Fig. 1). Moreover, such piecemeal networks would cost far more to establish and maintain, and presumably lose many more of their species through isolation and edge effects. Studies of both area-based and random selection therefore confirm that gathering at least some biological information is well worthwhile.

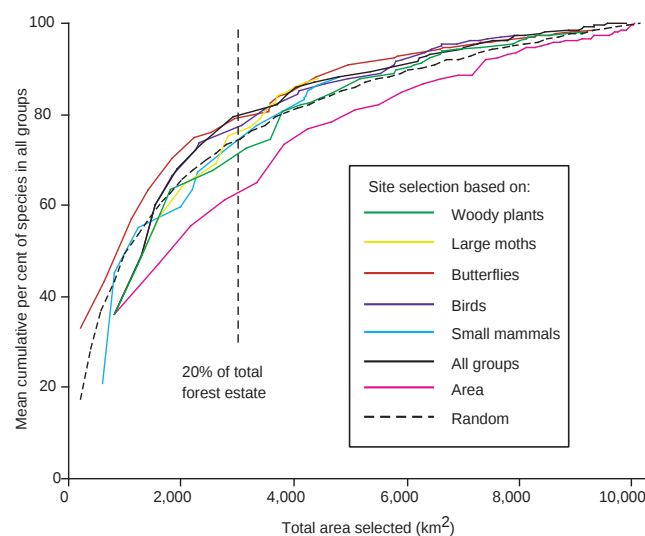


Figure 1 Mean cumulative representation of species across all five groups (woody plants, large moths, butterflies, birds and small mammals) as a function of cumulative area, when sites are picked on the basis of data on single taxa, on all groups, on forest area, or at random. Relatively few forests were picked using data on large moths or small mammals because all of the relatively few species in these groups could be captured in quite small networks of sites. The 20% threshold is drawn for illustrative purposes only. In reality, the Forest Department selected parts of (as well as whole) forests¹⁵, but, for simplicity, we compare algorithm performance when only whole reserves are chosen.

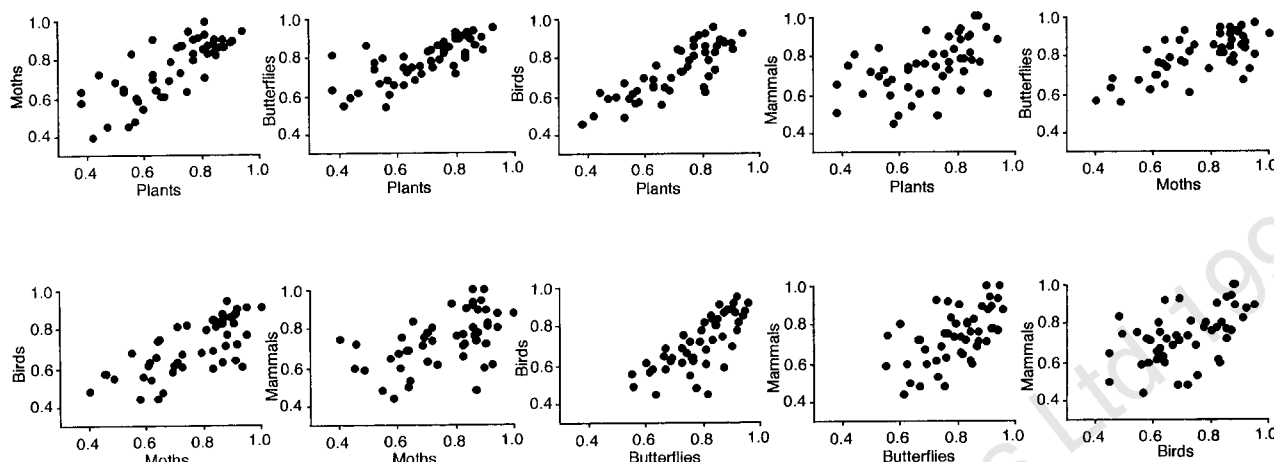


Figure 2 Cross-taxon congruence in complementarity scores²⁷ between 49 pairs of forests.

Why, in this case, are data on butterflies or birds alone just as useful for priority-setting as information on an entire suite of taxa, given that congruence in species richness is so strikingly low? The answer lies in the key issue of complementarity. The overall efficiency of a network of sites depends not just on their species richness but also on how well they complement one another biologically^{9,10}. In the same way, the extent to which data on a single taxon can be used to establish an efficient network of sites for conserving all groups depends on the extent not just of congruence in species richness, but also of cross-taxon congruence in patterns of complementarity. This point has been recognized only recently^{1,24–26}, and remains almost completely untested.

We studied this overlooked yet critical issue directly by calculating standardized complementarity scores²⁷ for pairs of forests; we then compared these scores across different groups. Cross-taxon congruence in these scores was consistently high (Fig. 2; Table 1). Pairs of forests with high complementarity for one taxon generally exhibited correspondingly high complementarity for all other taxa, and *vice versa*. This was not due to any confounding effect of differences between forests in sampling effort or area (see Methods).

Instead, cross-taxon congruence in complementarity arose because all the groups examined showed fundamentally similar patterns of biogeography. Modified TWINSpan analyses²⁸ of the biogeographical affinities of the forest reserves consistently identified six major groups (montane forests, western mid-altitude forests, lakeshore forests, savanna/forest mosaics, moist savannas and dry northern savannas), regardless of which taxon's distribution was assessed. Forests of different types exhibited high complementarity. Thus the top sites picked in any run of the priority-setting algorithm typically represented several forest types, and so collectively contained a broad set of species from each of our focal taxa.

These results, from the most detailed assessment of the performance of indicators for reserve selection anywhere in the tropics, show that in Uganda cross-taxon congruence in biogeography (and hence in between-site complementarity) is high enough to overwhelm low congruence in species richness. As a result, some groups can, by themselves, reliably indicate the overall conservation importance of forests. Thus similar priority-setting surveys in the region could in future be carried out more cheaply, without compromising their value, by focusing on just one or two taxa. As was the case here, exactly which groups perform best as indicator taxa will probably depend as much on the availability of local expertise as on the biology of the groups.

Our results also indicate that previous assessments of cross-taxon surrogacy in priority-setting, based on quantifying congruence in

species richness, may have been inadequate. Certain groups may function as good indicators for reserve selection in other areas that, like Uganda, are characterized by high biogeographical heterogeneity. More appropriate analyses of putative indicators that examine congruence in complementarity and assess the overall value of the networks of sites identified by those groups are now needed for other parts of the tropics. □

Methods

Surveys. Surveys were carried out by teams of biologists and specially trained forest rangers using established techniques (described in ref. 14). Forest reserves differed considerably in area; to ensure that roughly constant fractions of the individuals in a forest were sampled, survey effort was spread in proportion to forest size (regressions of number of days of inventory versus forest area were significant at $P < 0.001$ for all five groups).

Reserve-selection algorithm. This algorithm selected first the forest with the most species, then that with the biggest complement of species (that is, the most species not represented in the selected site), and so on, until all species were represented at least once. In the event of ties, the smaller of the tied forests was picked. By incorporating both between-site complementarity and species richness, this sort of algorithm has higher efficiency (that is, it captures more species in a given area⁹) than site selection based solely on species richness^{9,10,29,30}. Moreover, for the type of problem tackled here, where we aimed to select as priorities a substantial minority of a relatively small number of sites, this 'simple greedy' algorithm has similar power to more sophisticated techniques^{11,24}. Our results deal with species accumulation as a function of total area of sites selected, but our conclusions were unchanged when we examined species accumulation versus number of sites selected.

Cross-taxon congruence in complementarity. Complementarity scores were calculated as the sum of the number of species found at just one or the other of a pair of sites, divided by the combined total found at either or both sites²⁷. To avoid cross-taxon correlations being inflated by pseudoreplication or driven by marked differences in area between paired forests, we calculated complementarity scores only between each site and the next smallest forest in our study (yielding $N = 49$ scores from 50 forests). High cross-taxon congruence in complementarity was not driven by potentially confounding effects of forest size or sampling effort: residual complementarity scores (controlling for significant relationships between complementarity and forest area, sampling effort, and the difference in these values between forests) remained highly intercorrelated across taxa ($P < 0.001$ for all cross-taxon correlations using residual values).

Received 31 December 1997; accepted 18 May 1998.

1. Kremen, C. *et al.* Terrestrial arthropod assemblages: their use in conservation planning. *Conserv. Biol.* **7**, 796–808 (1993).
2. Flather, C. H., Wilson, K. R., Dean, D. J. & McComb, W. C. Identifying gaps in conservation networks: of indicators and uncertainty in geographic-based analyses. *Ecol. Appl.* **7**, 531–542 (1997).

3. Prendergast, J. R., Quinn, R. M., Lawton, J. H., Eversham, B. C. & Gibbons, D. W. Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* **365**, 335–337 (1993).
4. Lombard, A. T. The problems with multi-species conservation: do hotspots, ideal reserves and existing reserves coincide? *S. Afr. J. Zool.* **30**, 145–163 (1995).
5. Gaston, K. J. Biodiversity—congruence. *Prog. Phys. Geog.* **20**, 105–112 (1996).
6. Gaston, K. J. in *Aspects of the Genesis and Maintenance of Biological Diversity* (eds Hochberg, M. E., Clobert, J. & Barbault, R.) 221–242 (Oxford Univ. Press, Oxford, 1996).
7. Oliver, I. & Beattie, A. J. Designing a cost-effective invertebrate survey: a test of methods for rapid assessment of biodiversity. *Ecol. Appl.* **6**, 594–607 (1996).
8. Kerr, J. T. Species richness, endemism, and the choice of areas for conservation. *Conserv. Biol.* **11**, 1094–1100 (1997).
9. Pressey, R. L. & Nicholls, A. O. Efficiency in conservation evaluation: scoring versus iterative approaches. *Biol. Conserv.* **50**, 199–218 (1989).
10. Pressey, R. L., Humphries, C. J., Margules, C. R., Vane-Wright, R. I. & Williams, P. H. Beyond opportunism: key principles for systematic reserve selection. *Trends Ecol. Evol.* **8**, 121–128 (1993).
11. Williams, P. H. et al. A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British birds. *Conserv. Biol.* **10**, 155–174 (1996).
12. Pomeroy, D. Centers of high biodiversity in Africa. *Conserv. Biol.* **7**, 901–907 (1993).
13. Howard, P. C. *Nature Conservation in Uganda's Tropical Forest Reserves*. (IUCN, Gland, 1991).
14. Howard, P. C. & Davenport, T. R. B. (eds) *Forest Biodiversity Reports Vols 1–33* (Uganda Forest Department, Kampala, 1996).
15. Howard, P. C., Davenport, T. R. B. & Kigenyi, F. Planning conservation areas in Uganda's natural forests. *Oryx* **31**, 253–264 (1997).
16. Pearson, D. L. & Cassola, F. World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): indicator taxon for biodiversity and conservation studies. *Conserv. Biol.* **6**, 376–391 (1992).
17. Schall, J. J. & Pianka, E. R. Geographical trends in numbers of species. *Science* **201**, 679–686 (1978).
18. Lawton, J. H. et al. Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* **391**, 72–76 (1998).
19. van Jaarsveld, A. S. et al. Biodiversity assessment and conservation strategies. *Science* **279**, 2106–2108 (1998).
20. Prendergast, J. R., Wood, S. N., Lawton, J. H. & Eversham, B. C. Correcting for variation in recording effort in analyses of diversity hotspots. *Biodiv. Lett.* **1**, 39–53 (1993).
21. Rytty, R. T. Effect of the focal taxon on the selection of nature reserves. *Ecol. Appl.* **2**, 404–410 (1992).
22. Sætersdal, M., Line, J. M. & Birks, H. J. B. How to maximize biological diversity in nature reserve selection: vascular plants and breeding birds in deciduous woodlands, western Norway. *Biol. Conserv.* **66**, 131–138 (1993).
23. Dobson, A. P., Rodriguez, J. P., Roberts, W. M. & Wilcove, D. S. Geographic distribution of endangered species in the United States. *Science* **275**, 550–553 (1997).
24. Casati, B. et al. A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon. *Biol. Conserv.* **80**, 83–97 (1997).
25. Faith, D. P. & Walker, P. A. How do indicator groups provide information about the relative biodiversity of different sets of areas? on hotspots, complementarity and pattern-based approaches. *Biodiv. Lett.* **3**, 18–25 (1996).
26. Oliver, I., Beattie, A. J. & York, A. Spatial fidelity of plant, vertebrate and invertebrate assemblages in multiple-use forest in eastern Australia. *Conserv. Biol.* **12**, (in the press).
27. Colwell, R. K. & Coddington, J. A. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. Lond. B* **345**, 101–108 (1994).
28. Oksanen, J. & Minchin, P. R. Instability of ordination results under changes in input data order: explanations and remedies. *J. Veg. Sci.* **8**, 447–454 (1997).
29. Kershaw, M., Williams, P. H. & Mace, G. M. Conservation of Afrotropical antelopes: consequences and efficiency of using different site selection methods and diversity criteria. *Biodiv. Conserv.* **3**, 354–372 (1994).
30. Williams, P. H. in *Conservation in a Changing World* (eds Mace, G. M., Balmford, A. & Ginsberg, J. R.) (Cambridge University Press, in the press).

Acknowledgements. We thank our colleagues in Uganda, especially members of the Forest Department field inventory teams, as well as R. Badaza, D. Duli, D. Hafashimana, I. Kapalaga, T. Katende, R. Kityo, E. Mupada, R. Nabanyumya, D. Olet, A. Rodgers, R. Murtland and T. Finch. We also thank P. Warren and J. Ward for help with analyses, and T. Birkhead, K. Gaston, J. Lawton, D. Pomeroy and P. Williams for comments on the manuscript. Most of the work was funded by the European Union and the Global Environmental Facility; P.V. was supported through the United Nations Food and Agriculture Organisation and J.S.L. and A.B. by the Darwin Initiative.

Correspondence and requests for materials should be addressed to A.B. (e-mail: a.balmford@sheffield.ac.uk).

Membrane potential synchrony of simultaneously recorded striatal spiny neurons *in vivo*

Edward A. Stern*, Dieter Jaeger† & Charles J. Wilson*

* Department of Anatomy and Neurobiology, University of Tennessee College of Medicine, 855 Monroe Avenue, Memphis, Tennessee 38163, USA

† Department of Biology, Emory University, Atlanta, Georgia 30322, USA

The basal ganglia are an interconnected set of subcortical regions whose established role in cognition and motor control remains poorly understood. An important nucleus within the basal ganglia, the striatum, receives cortical afferents that convey sensorimotor, limbic and cognitive information¹. The activity of

medium-sized spiny neurons in the striatum seems to depend on convergent input within these information channels². To determine the degree of correlated input, both below and at threshold for the generation of action potentials, we recorded intracellularly from pairs of spiny neurons *in vivo*. Here we report that the transitions between depolarized and hyperpolarized states were highly correlated among neurons. Within individual depolarized states, some significant synchronous fluctuations in membrane potential occurred, but action potentials were not synchronized. Therefore, although the mean afferent signal across fibres is highly correlated among striatal neurons, the moment-to-moment variations around the mean, which determine the timing of action potentials, are not. We propose that the precisely timed, synchronous component of the membrane potential signals activation of cell assemblies and enables firing to occur. The asynchronous component, with low redundancy, determines the fine temporal pattern of spikes.

The membrane potential of striatal spiny projection neurons recorded in anaesthetized animals *in vivo* fluctuates between two subthreshold states^{3–5}. The quiescent, hyperpolarized 'down' state, and the noisier, depolarized 'up' state, are separated by 15–30 mV. Spike threshold is usually 3–5 mV above the mean potential of the 'up' state⁶. The 'up' state is caused by synchronous synaptic input from a large number of corticostriatal afferents interacting with nonlinear membrane conductances in the striatal spiny neurons⁴. Because individual corticostriatal synapses are relatively weak⁷, many afferents must be activated to cause a striatal neuron to fire⁸. Neurons within functionally related striatal microzones receive similar sets of afferents, which arise from spatially distributed, but functionally related, cortical areas^{9,10}. Although the neurons within a microzone respond to behavioural events in a time-locked manner¹¹, their responses are asynchronous¹². We recorded intracellularly from 23 pairs of striatal neurons in rats anaesthetized with urethane and ketamine/xylazine to measure the amount of synchrony of the shared corticostriatal afferents.

The membrane potentials of all the neurons in our sample showed distinct 'up' and 'down' states. Membrane potential traces of a pair of simultaneously recorded spiny neurons are shown in Fig. 1a. All-points histograms of 10-s samples demonstrate the bimodality of the membrane potentials (Fig. 1b). Cross-correlograms of the times of the transitions from the 'down' state to the 'up' state ('up' transitions) and transitions from the 'up' state to the 'down' state ('down' transitions) were constructed to measure the synchrony of the subthreshold membrane-potential state transitions (Fig. 2). All cross-correlograms showed significant central peaks (outside the 95% confidence intervals of the correlogram), indicating synchrony of both 'up' and 'down' state transitions. No further significant peaks were found in our correlograms, indicating that the state transitions are not due to a single oscillatory process⁵.

We tested the hypothesis that synchrony of the state transitions of spiny neurons is dependent on the distance between them. The synchrony of the state transitions was weakly anticorrelated with distance ($r = -0.40$, $t = 3.14$, $P < 0.05$). The correlations found between the neurons in all the pairs of our sample showed that a large proportion of the state transitions of the striatal population are synchronized even over distances of a millimetre or more.

We calculated cross-correlations of the membrane potentials within individual 'up' states of simultaneously recorded spiny cells to see whether correlations on the time scale of individual synaptic inputs could be detected. The membrane potentials within the 'up' states of the spiny neurons were variable, reflecting the incoming corticostriatal synaptic activity⁵. The cross-correlations of the individual 'up' states were also variable. For each pair, we averaged the cross-correlations over several 'up' states. In 30% (7 of 23 cells) of our sample, a large central peak (outside the 95% confidence limits^{13,14}) was evident. The central peaks had half-widths varying from 5 to 10 ms, which is the same time scale as corticostriatal

excitatory postsynaptic potentials¹⁵. We found no significant correlation of the width, amplitude or phase of the peak with distance between the neurons. An example of the fluctuations within individual 'up' states of a pair of spiny neurons is shown in Fig. 3a (note the different time and voltage scales from Fig. 1a). Although the membrane potentials are within a single 'up' state, some correlation is apparent (shown by the grey bars). The correlation is on the time scale of tens of milliseconds, but is not readily apparent on the millisecond timescale. An example of the cross-correlations of the 'up' states is shown in Fig. 3b.

The peak in Fig. 3b is only slightly outside the 95% confidence limits. The average cross-correlation shown in Fig. 3b is above zero, which clearly contributed to the significance of the peak. The broad correlation, on the scale of tens of milliseconds, was present in the average 'up' states of all of our cell pairs. We performed shuffled cross-correlations of 'up' states from one cell with subsequent 'up' states from the other cell in the pair to see if the positive correlations of simultaneously recorded 'up' states are artefacts of our method. The average of the cross-correlations was not significantly different from zero, unlike the cross-correlations of the simultaneous 'up' states (Fig. 3d). No significant central peak was seen in any of

our shuffled correlations. Unlike the cross-correlations of the simultaneously recorded 'up' states, the average correlation was not significantly different from zero. The average positive correlogram in Fig. 3b may reflect the same global cortical correlation as the correlations of the state transitions, or it may be caused by a subset of the correlated inputs responsible for the global synchrony.

For comparison with the conventional approach for studying neuronal synchrony, we calculated the autocorrelograms and cross-correlograms of action potentials of the spiny cell pairs. Five pairs fired enough spikes for us to calculate sufficient statistics for the correlograms. The typical cross-correlogram showed a weak periodicity of ~ 1 Hz, corresponding to the average period of the state transitions. In all cases, the width of the peaks was representative of the mean duration of the 'up' states of the neurons; an example is shown in Fig. 4. In no case did the correlograms have the structure expected from direct synaptic connections between the neurons in the pair¹⁶. No significant narrow (~ 5 ms) peaks or troughs were observed in our spike cross-correlations. The Fourier transform of the cross-correlograms did not reveal significant peaks around ~ 1 Hz, showing that the structure of the spike correlograms represented only the increased probability of spiking resulting from the state of the membrane potential.

Spiny neurons have been shown to be dye-coupled under some circumstances¹⁷. Intracellular recording allowed us directly to test the hypothesis that spiny neurons are functionally connected. Depolarizing and hyperpolarizing current pulses passed in the 'up' and 'down' states failed to elicit any electrical responses in the

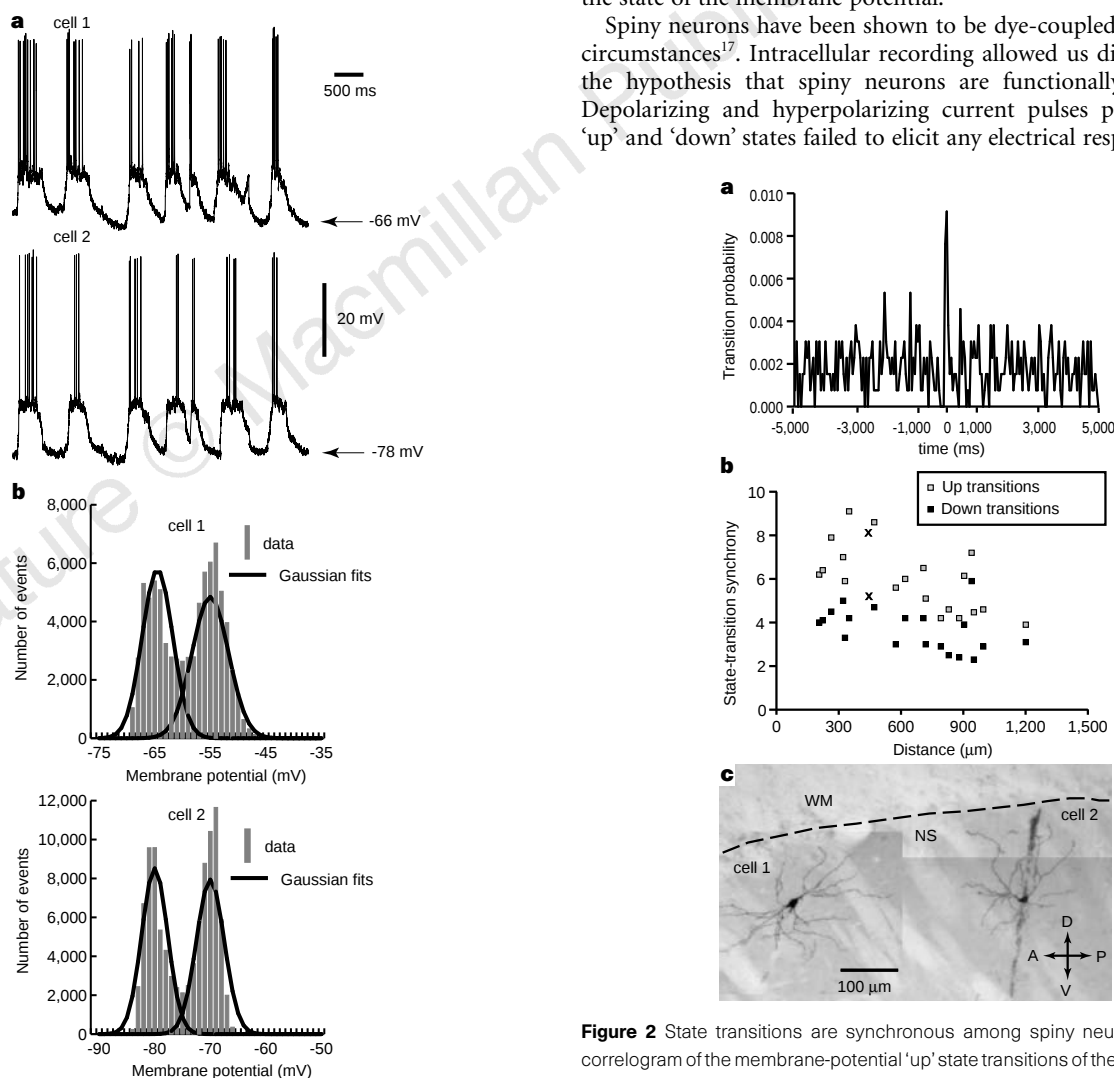


Figure 2 State transitions are synchronous among spiny neurons. **a**, Cross-correlogram of the membrane-potential 'up' state transitions of the cell pair shown in Fig. 1. **b**, Scatterplot showing state-transition synchrony (probability of first bin/standard deviation) as a function of distance between the neurons in the pair. Crosses denote the synchrony of the pair whose cross-correlogram is shown in **a**. **c**, Photomontage showing the striatal neurons from which the traces in **a** were recorded. The electrode track running through cell 2 is clearly seen. The broken line denotes the boundary between the white matter (WM) and the neostriatum (NS).

Figure 1 Membrane potentials of striatal spiny neurons show distinct 'up' and 'down' states. **a**, Simultaneous dual intracellular recording from a pair of striatal spiny neurons *in vivo*. Note the coherence of the 'up' and 'down' state transitions. Scale bars are shown on the right. **b**, All-points histograms of 10 s of the membrane-potential recordings shown in **a** (grey bars) and the two best-fitting gaussians best describing the data (black lines).

other cell in the pair. Averaging the membrane potential triggered by spontaneous and intracellularly evoked action potentials did not reveal any synaptic interactions between the neurons. Our results are consistent with similar dual intracellular recording studies in striatal slice preparations¹⁸ in which no direct physiological interactions between spiny cells were detected. We conclude that direct synaptic or electrical interactions between spiny neurons did not cause the synchronization of the membrane-potential state transitions.

We previously proposed that the membrane-potential state transitions of striatal neurons are caused by similar state transitions in a large population of corticostriatal afferents^{4,5}. Our present data show that coherence of the cortical afferents can synchronize the onsets of the 'up' states of striatal neurons, whereas the action potentials of the neurons remain largely asynchronous.

Each striatal neuron receives a large number, although a small

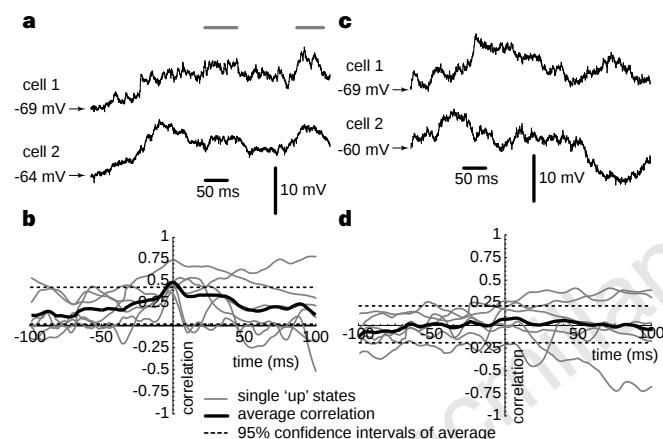


Figure 3 Membrane potential synchrony within 'up' states. **a**, Portions of an individual 'up' state in two simultaneously recorded spiny neurons. The mean membrane potential in the 'down' state was -80 mV for cell 1 and -75 mV for cell 2. **b**, Cross-correlation of the waveforms within individual simultaneously recorded 'up' states from a single pair of neurons. **c**, As **a**, but the 'up' state in cell 1 is the 'up' state subsequent to that from cell 0. **d**, Shuffled cross-correlation of the waveforms within individual 'up' states. The 'up' states of cell 2 were subsequent to those recorded from cell 1.

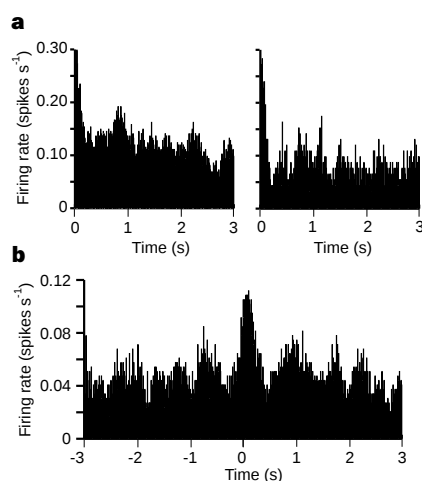


Figure 4 Spike auto- and cross-correlograms for a pair of striatal spiny neurons, calculated with 10 ms bin-width. **a**, The autocorrelograms show ~ 1 Hz broad periodicity, reflecting the average and variation of the 'up' state durations. **b**, The cross-correlogram shows no additional structure, apart from a change in firing rate of one of the neurons.

proportion, of the corticostriatal afferent population². The transitions between 'up' and 'down' states of striatal neurons in a specific functionally defined region will be synchronous because they depend on the total number of active excitatory inputs. The moment-to-moment variations of membrane potential are generally not synchronous on the time scale of a few milliseconds, as can be seen in Fig. 3b. This could result from activation of common input fibres projecting to both striatal neurons, or from synchronization of cortical or thalamic afferent neurons.

The subthreshold membrane-potential state transitions in cortical cells are correlated with the slow (~ 1 Hz) rhythm of the electroencephalogram (EEG) during sleep, and have been shown to be dependent on the level of anaesthesia^{19–21}. As the level of anaesthesia becomes more shallow, the state transitions among the cortical neurons become less synchronized, and the slow EEG rhythm becomes progressively more shallow until it is lost. However, it is clear that the state transitions in individual cortical neurons are still present²¹. Cortical state transitions have also been observed in animals anaesthetized with halothane and nitrous oxide²².

Membrane-potential state transitions in striatal spiny neurons are also not dependent on anaesthesia, as they were first reported in experiments on awake^{3,23} and decerebrate²⁴ animals. Similar striatal and cortical membrane-potential state transitions have been reported in organotypic cocultures of cortex, striatum and substantia nigra, demonstrating the robustness of this type of activity²⁵. In awake rats, the global synchrony seen in anaesthesia is replaced by functionally organized coherence on a more local scale. During voluntary movements, the firing of striatal neurons within micro-zones is time locked to the movements but is asynchronous¹². We propose that the response during the movements is associated with 'up' state transitions in the responsive cell assembly. The asynchrony of firing is caused by fluctuations of membrane potential within the 'up' state, which represents smaller subsets of afferent activity. The asynchrony reflects the low redundancy of the activity within the striatal cell assembly, which results in higher information capacity.

We have previously shown that the fluctuations of the membrane potential around the mean value of the 'up' states have a spectral composition that is inconsistent with shot noise driven synaptic conductances⁵. We propose that the fluctuations represent activation of small groups of related cortical inputs embedded within the total cortical input signal. In many parts of the mammalian nervous system, deviations from average or population values have been shown to carry essential signals^{26–29}. We suggest that a similar coding mechanism is operating in corticostriatal pyramidal and striatal spiny neurons. The precisely timed, synchronous transitions to the 'up' state may signal the general activation of a given microzone by a behavioural event and enable firing. More specific information may then be carried by the fine structure of the asynchronous spike trains of the neurons within the assembly.

Methods

General. All the procedures were carried out in accordance with animal care guidelines of the US National Institutes of Health and the University of Tennessee.

Surgery. The experiments were performed on 15 male Sprague-Dawley or Long-Evans rats weighing 268–520 g, initially anaesthetized with urethane (1.25 g per kg body weight, intraperitoneal), and given hourly supplemental injections of ketamine–xylazine (30 mg per kg and 7 mg per kg, respectively, intramuscular). The animals' temperatures were maintained at 37°C . The animals were placed in a stereotaxic instrument and the skull exposed with a single incision. The cisterna magna was opened to relieve intracranial pressure. All animals continued to breathe without artificial respiration, and were suspended by a clamp at the base of the tail to minimize movements of the brain caused by breathing. For insertion of the recording electrodes, a large opening was made in the skull from the frontal pole of the cortex to 3 mm posterior to bregma.

Electrophysiology and staining. Recording electrodes were glass micropipettes filled with 2–4% biocytin (Sigma) dissolved in 1 M potassium acetate. Electrode resistances ranged from 30 to 100 MΩ. The recording electrodes were aligned so that the tips would meet approximately 1–2 mm ventral to the dorsal surface of the striatum. After recording electrodes were inserted, the exposed cortex was covered with a low-melting-point paraffin wax to reduce brain pulsations. Recordings were made using an active bridge amplifier and then filtered and digitized at a rate of 4 kHz or higher. Neurons that had membrane potentials more negative than –60 mV and action potentials more positive than 0 mV were included in the sample. Neurons were stained by passing 1.5-nA depolarizing pulses of 300 ms duration every 600 ms for 10–60 min.

Histology. At the end of the experiment, the animals were given a lethal dose of Nembutal or urethane intraperitoneally and were perfused intracardially with isotonic buffered saline followed by 400 ml of 4% formaldehyde in 0.15 M sodium phosphate buffer (pH 7.2–7.6). Intact brains were removed and postfixed in cold phosphate-buffered formaldehyde. The brains were trimmed and cut on a vibratome in 50-μm parasagittal sections. Sections containing labelled neurons or processes were processed for biocytin³⁰. Some sections were postfixed with 1% osmium tetroxide. Cells were identified by the coordinates of the recording electrodes, which were recorded during the experiment. Only pairs in which both neurons were identified were used in the sample.

Data analysis. The membrane potential distribution was normalized and the weighted sum of two gaussians was fit to the data using the Levenberg–Marquart method of nonlinear least squares. The model used was

$$p(v) = \frac{a}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(v-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1-a}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(v-\mu_2)^2}{2\sigma_2^2}\right)$$

where μ_1 is the mean membrane potential in the ‘down’ state; σ_1 is the amplitude of fluctuations within the ‘down’ state; μ_2 is the mean membrane potential in the ‘up’ state; σ_2 is the amplitude of fluctuations within the ‘up’ state and a is the ratio of ‘down’ state to ‘up’ state. After the two best-fitting gaussians were found, we defined the thresholds of the ‘up’ and ‘down’ states as half of the distance between the minimum point between the two gaussians and the modal point for that state. Using these methods, the state transition times were converted into a point process, which was used to construct the auto- and cross-correlograms of the state transitions. Cross-correlograms were constructed from the transition times of the two states, rather than cross-correlating the waveforms of the membrane potential, so that the only contribution to the central peak would be the jitter in the delay between the two neurons’ membrane-potential transition times. Cross-correlating the membrane-potential waveforms would result in a broader peak owing to the variability of the autocorrelation waveforms, that is, the durations of the ‘up’ and ‘down’ states in each cell. Auto- and cross-correlograms of state transitions and spikes were calculated using standard methods¹⁶ implemented using the program Mathematica. All cross-correlograms were compared to autocorrelograms of the component neurons. The central peak in the cross-correlogram, which was present for each cell pair, was never present in the autocorrelograms. Synchrony (strength of cross-correlation) was calculated by dividing the magnitude of the central peak by the standard deviation of the correlogram. Cross-correlation of the waveforms was performed using standard time-series methods¹³ implemented in Mathematica. The cross-correlations were calculated starting from 100 ms from the first peak within the ‘up’ state. Only ‘up’ states without spikes were selected. 95% confidence intervals were calculated for the harmonic means of 6 cross-correlations.

Received 9 March; accepted 8 June 1998.

- Graybiel, A. M., Aosaki, T., Flaherty, A. W. & Kimura, M. The basal ganglia and adaptive motor control. *Science* **265**, 1826–1831 (1994).
- Wilson, C. J. in *The Synaptic Organization of the Brain* (ed. Shepherd, G.) 329–375 (Oxford Univ. Press, New York, 1997).
- Wilson, C. J. & Groves, P. M. Spontaneous firing patterns of identified spiny neurons in the rat neostriatum. *Brain Res.* **220**, 67–80 (1981).
- Wilson, C. J. & Kawaguchi, Y. The origins of two-state spontaneous membrane potential fluctuations of neostriatal spiny neurons. *J. Neurosci.* **16**, 2397–2410 (1996).
- Stern, E. A., Kincaid, A. E. & Wilson, C. J. Spontaneous subthreshold membrane potential fluctuations and action potential variability of rat corticostriatal and striatal neurons in vivo. *J. Neurophysiol.* **77**, 1697–1715 (1997).
- Wickens, J. R. & Wilson, C. J. Regulation of action potential firing in spiny neurons of the rat neostriatum in vivo. *J. Neurophysiol.* **79**, 2358–2364 (1998).
- Choi, S. & Lovinger, D. M. Decreased frequency but not amplitude of quantal synaptic responses associated with expression of corticostriatal long-term depression. *J. Neurosci.* **17**, 8613–8620 (1997).
- Wilson, C. J. in *Single Neuron Computation* (eds McKenna, T., Davis, J. & Zornetzer, S. F.) 141–171 (Academic, San Diego, 1992).

- Alexander, G. E. & DeLong, M. R. Microstimulation of the primate neostriatum. I. Physiological properties of striatal microexcitable zones. *J. Neurophysiol.* **53**, 1401–1416 (1985).
- Flaherty, A. W. & Graybiel, A. M. Two input systems for body representations in the primate striatal matrix: Experimental evidence in the squirrel monkey. *J. Neurosci.* **13**, 1120–1137 (1993).
- Alexander, G. E. & DeLong, M. R. Microstimulation of the primate neostriatum. II. Somatotopic organization of striatal microexcitable zones and their relation to neuronal response properties. *J. Neurophysiol.* **53**, 1417–1430 (1985).
- Jaeger, D., Gilman, S. & Aldridge, J. W. Neuronal activity in the striatum and pallidum of primates related to the execution of externally cued reaching movements. *Brain Res.* **694**, 11–127 (1995).
- Diggle, P. *Time Series* (Oxford Univ. Press, 1990).
- Abeles, M. Quantification, smoothing, and confidence limits for single-units’ histograms. *J. Neurosci. Methods* **5**, 317–325 (1982).
- Wilson, C. J. Postsynaptic potentials evoked in spiny neostriatal projection neurons by stimulation of ipsilateral and contralateral cortex. *Brain Res.* **367**, 201–213 (1986).
- Perkel, D. H., Gerstein, G. L. & Moore, G. P. Neuronal spike trains and stochastic point processes. *Biophys. J.* **7**, 419–440 (1967).
- Onn, S. P. & Grace, A. A. Dye coupling between rat striatal neurons recorded in vivo; compartmental organization and modulation by dopamine. *J. Neurophysiol.* **71**, 1917–1934 (1994).
- Jaeger, D., Kita, H. & Wilson, C. J. Surround inhibition among projection neurons in weak or nonexistent in the rat neostriatum. *J. Neurophysiol.* **72**, 2555–2558 (1994).
- Steriade, M., Nunez, A. & Amzica, F. A novel slow (<1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *J. Neurosci.* **13**, 3252–3265 (1993).
- Amzica, F. & Steriade, M. Short- and long-range neuronal synchronization of the slow (<1 Hz) cortical oscillation. *J. Neurophysiol.* **73**, 20–38 (1995).
- Contreras, D. & Steriade, M. State-dependent fluctuations of low-frequency rhythms in corticothalamic rhythms. *Neuroscience* **76**, 25–38 (1995).
- Douglas, R. J., Martin, K. A. C. & Whitteridge, D. An intracellular analysis of the visual responses of neurons in cat visual cortex. *J. Physiol. (Lond.)* **440**, 659–696 (1991).
- Wilson, C. J. The generation of natural firing patterns in neostriatal neurons. *Prog. Brain Res.* **99**, 277–297 (1993).
- Hull, C. D., Bernardi, G. & Buchwald, N. A. Intracellular responses of caudate neurons to brain stem stimulation. *Brain Res.* **22**, 163 (1970).
- Plenz, D. & Kitai, S. T. Up and down states in striatal medium spiny neurons simultaneously recorded with spontaneous activity in fast-spiking interneurons studied in cortex-striatum-substantia nigra organotypic cultures. *J. Neurosci.* **18**, 266–283 (1998).
- Stern, E. A., Aertsen, A., Vaadia, E. & Hochstein, S. Stimulus encoding by multidimensional receptive fields in single cells and cell populations in V1 of awake monkey. *Adv. Neural Inf. Process. Systems* **5**, 377 (1993).
- Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* **273**, 1868–1871 (1996).
- O’Keefe, J. & Reece, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
- Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
- Horikawa, K. & Armstrong, W. E. A versatile means of intracellular labeling: injection of biocytin and its detection with avidin conjugates. *J. Neurosci. Methods* **25**, 1–11 (1988).

Acknowledgements. This work was supported by a grant from the NIH. We thank B. Mattix for technical assistance.

Correspondence and requests for materials should be addressed to E.S. (edstern@marlin.utm.edu).

Glutamate locally activates dendritic outputs of thalamic interneurons

Charles L. Cox, Qiang Zhou & S. Murray Sherman

Department of Neurobiology, State University of New York, Stony Brook, New York 11794-5230, USA

The relay of information through thalamus to cortex is dynamically gated, as illustrated by the retinogeniculocortical pathway¹. Important to this is the inhibitory interneuron in the lateral geniculate nucleus (LGN). For the typical neuron, synaptic information arrives through postsynaptic dendrites and is transmitted by axon terminals. However, the typical thalamic interneuron, in addition to conventional axonal outputs, has distal dendrites that serve both pre- and postsynaptic roles^{2–6}. These dendritic terminals participate in curious and enigmatic triadic arrangements, in which each contacts a relay cell dendrite and is contacted by a glutamatergic retinal terminal that innervates the same relay cell dendrite. Here we show that agonists of the metabotropic glutamate receptor (mGluR) activate dendritic terminals of interneurons in the absence of action potentials, thereby inhibiting the postsynaptic relay neuron. Somatic recordings from LGN interneurons reveal that there is no response to mGluR agonists, suggesting that their dendritic terminals are electrically isolated

from their somata and axons, consistent with anatomical modeling of these cells⁷. Our results offer insight into the functioning of triadic circuitry and indicate that thalamic interneurons can perform independent computations expressed through axonal as opposed to dendritic outputs.

Local, GABAergic, inhibitory inputs to dendrites of relay cells of the LGN derive from three sources: two are axonal—from interneurons and neurons in the thalamic reticular nucleus (TRN)—and the other is from the dendritic terminals of interneurons (see also below). The axonal terminals, also known as F1 terminals, are presynaptic only, and form conventional inhibitory synaptic contacts. The dendrites of the interneurons have been described as “axoniform”⁸, because at their distal regions they give rise to small, thin stalks terminating in dendritic terminals. These are known as F2 terminals, and they are unusual in that they are both pre- and postsynaptic (reviewed in ref. 1; Fig. 1A). These F2 terminals are usually involved in triads (Fig. 1A, shaded area), in which the retinogeniculate terminal provides excitatory contacts onto both a relay cell dendrite and the F2 terminal, with the F2 terminal contacting the same relay cell dendrite (Fig. 1A). This arrangement suggests that the retinal input produces both monosynaptic excitation and feedforward disinhibition at the relay cell. Finally, although all LGN relay cells receive inputs from F1 terminals, only morphological type 2 cells (thought to be X; see refs 9, 10) receive significant numbers of F2 inputs¹¹, and type 1 cells (thought to be Y) do not; thus, only type 2 cells should display physiology reflecting triadic inputs involving F2 terminals¹¹. The problem has been to test the function of these triadic circuits, which have previously been defined anatomically but not physiologically because of the difficulty in recording from or selectively activating the F2 terminals.

A possible solution to this problem was offered by immunocytochemical studies of the distribution of mGluRs in the cat's LGN¹², which indicated that mGluRs are selectively located postsynaptic to cortical inputs on relay cell dendrites and on F2 terminals postsynaptic to retinal input. By suppressing the postsynaptic actions of mGluR activation in the relay neurons, this arrangement allows the possibility of selectively stimulating F2 terminals with mGluR

agonists, thereby gaining a glimpse into the functioning of these terminals specifically and interneurons more generally. We did this by using whole-cell recording of relay cells and interneurons from *in vitro* thalamic slice preparations of cats and rats.

During current clamp recording of cat LGN relay cells, the general mGluR agonist ($\pm$)-1-aminocyclopentane-*trans*-1,3-dicarboxylic acid (ACPD) produced a long-lasting membrane depolarization associated with an increased input resistance (Fig. 1B). This presumably acts through the mGluRs on relay cell dendrites and is consistent with previous demonstrations that mGluR activation reduces a resting potassium ‘leak’ conductance (K_{leak}), thereby depolarizing the membrane^{13,14}. We also noticed that ACPD application produced a large increase in the baseline activity (Fig. 1Bb) that persisted independently of the ACPD-mediated membrane depolarization (Fig. 1Bc) and was suppressed by the GABA_A antagonist bicuculline methiodide (BMI; data not shown). We conclude that this increased activity reflected spontaneous inhibitory postsynaptic potentials (sIPSPs), which also implies that this effect is independent of the ACPD-mediated reduction of K_{leak} in the postsynaptic neuron.

To help isolate spontaneous inhibitory postsynaptic currents (sIPSCs), we minimized the effect of ACPD on K_{leak} by substituting Cs⁺ for K⁺ in the recording pipette (Cs⁺ will thus diffuse into the cell and block the K⁺ conductance underlying K_{leak}), and we made voltage clamp recordings with a holding potential of 0 mV to increase the driving force of the presumed Cl[−] conductance (Fig. 1C, D). Under these conditions, application of ACPD (125–250 μ M) increased the amplitude and frequency of sIPSCs ($n = 19$ cells; Fig. 1C). These sIPSCs were eliminated or strongly attenuated by BMI (30 μ M; $n = 8$ of 8 cells tested; Fig. 2Ad), and their reversal potential was near the calculated Cl[−] equilibrium potential (data not shown). This further supports the conclusion that these sIPSCs activated by ACPD application represented activation of GABA_A receptors.

One explanation for the increased sIPSC activity is that ACPD activated local inhibitory neurons (TRN cells and/or interneurons) that form synaptic contacts through F1 terminals onto the recorded neurons (Fig. 3A). To eliminate this possibility, we applied

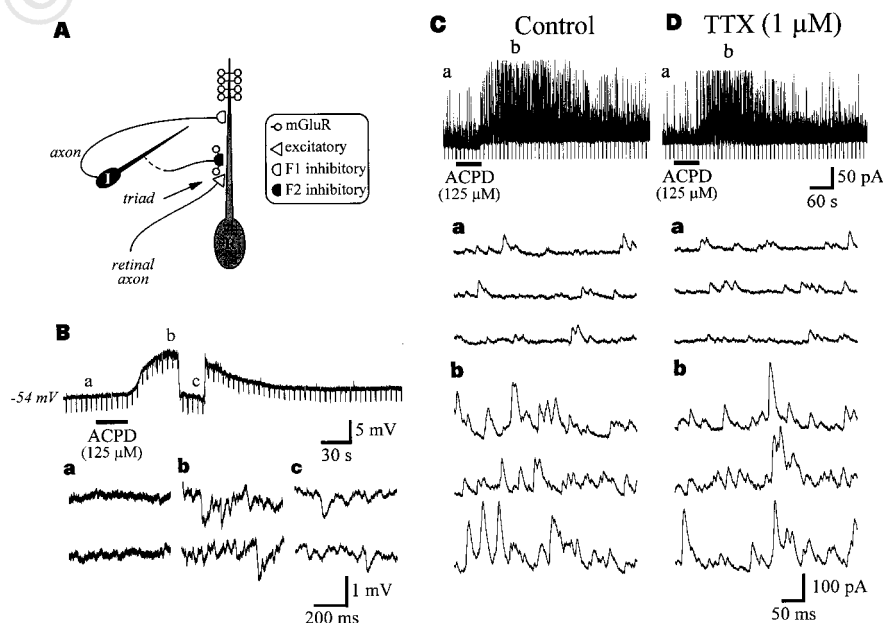


Figure 1 Activation of mGluRs increases inhibition in cat LGN relay neurons. **A**, The retinal axon forms excitatory synapses onto both the relay cell (R) dendrite and the interneuron's (I) F2 terminals. F2 terminals emanate from distal dendrites after many branches (dashed portion) and synapse onto the relay cell. **B**, Current clamp recording from relay cell. ACPD produces a long-lasting depolarization and an increase in baseline activity. The expanded traces (a, b, c) show the baseline

activity during the indicated periods. Downward deflections are responses to constant current pulses used to estimate input resistance. **C**, **D**, Voltage clamp recording from relay neuron, **C**, ACPD produces a long-lasting increase in sIPSC activity, as shown in a, b and then expanded traces below. **D**, This ACPD-mediated increase in sIPSC activity persists in TTX, as shown in a, b.

tetrodotoxin (TTX) to prevent these inhibitory cells from producing action potentials that could be conducted down their axons to inhibit the recorded relay cells. In 7 of 19 neurons tested, TTX eliminated the ACPD-induced increase in sIPSC activity, suggesting that in these 7 cells the evoked inhibition could be completely explained on the basis of ACPD activation of TRN cells or interneurons. Previous evidence¹⁵, which we confirm here, indicates that interneurons do not respond with increased firing rates to ACPD. However, TRN cells appear to have abundant mGluRs¹² and ACPD application strongly increases their firing rate^{16,17}. Thus, the TTX-dependent increase in IPSCs we observed for the 7 cells probably involved an induced increase in activity of TRN inputs to these relay cells.

Surprisingly, the robust increase in sIPSC activity persisted in the presence of TTX in the remaining 12 relay cells (Fig. 1D). In most of these cells, the ACPD effect was slightly attenuated by TTX, suggesting that these neurons contained a TTX-sensitive component of this increased sIPSC activity. As already noted, TTX application eliminates inputs to the relay cell from axonal F1 terminals, which require action potentials for activation, but presumably has little effect on F2 terminals, implying that they do not require action potentials to become activated. If so, then 7 of the recorded cells received synaptic inputs from F1 terminals but from virtually no F2 terminals, as expected for type 1 cells (see above); the remaining 12 cells received inputs from both F1 and F2 terminals, as expected for type 2 cells.

To determine whether the remaining, TTX-insensitive, increase in sIPSC activity resulted from mGluR-evoked GABA release from presynaptic terminals (for example, the F2 terminals) or by an as-yet

unidentified postsynaptic effect, we repeated some of these experiments in a low Ca^{2+} (0.2 mM) and high Mg^{2+} (6 mM) extracellular solution. This should eliminate sIPSCs evoked by synaptic transmission, which is Ca^{2+} -dependent. Under these conditions, the TTX-insensitive ACPD-mediated effect was completely suppressed (Fig. 2Aa–c). As an additional control, the amplitudes of the Ca^{2+} -independent miniature IPSCs were not altered by ACPD, indicating the lack of a postsynaptic site of the mGluR-mediated effect in the relay neuron. We conclude that mGluR activation leads to increased GABA release from synaptic terminals, resulting in increased activity of sIPSCs in the recorded relay cells.

From these data alone, we could not determine whether ACPD activates mGluRs directly on axonal (F1) or dendritic (F2) terminals, or both. We therefore took advantage of the unique anatomical organization of the rat thalamus (Fig. 3A). Although the LGN of the rat contains many interneurons and F2 terminals (like that of the cat), the ventrobasal complex (VB) is essentially devoid of

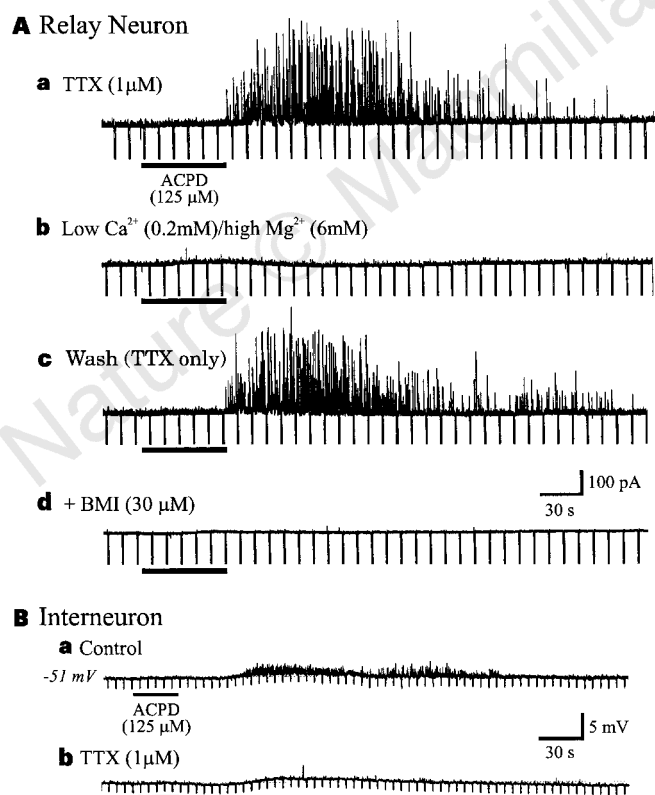


Figure 2 Voltage and current clamp recordings from cat LGN neuron. **A**, Voltage and current clamp recordings from cat LGN relay neuron demonstrate the synaptic nature of ACPD-mediated increase in sIPSCs. **a**, In TTX, ACPD increases sIPSCs; **b**, in a low- Ca^{2+} /high- Mg^{2+} solution, ACPD produces no change in sIPSC activity; **c**, the ACPD-mediated alteration in sIPSCs recovers following a 20 min wash; **d**, BMI attenuates the sIPSC activity, with and without ACPD. **B**, Current clamp recording from LGN interneuron. **a**, ACPD produces a small, long-latency depolarization (<1 mV) with an increase in sEPSPs. **b**, TTX blocks the increased sEPSPs, but the small depolarization persists.

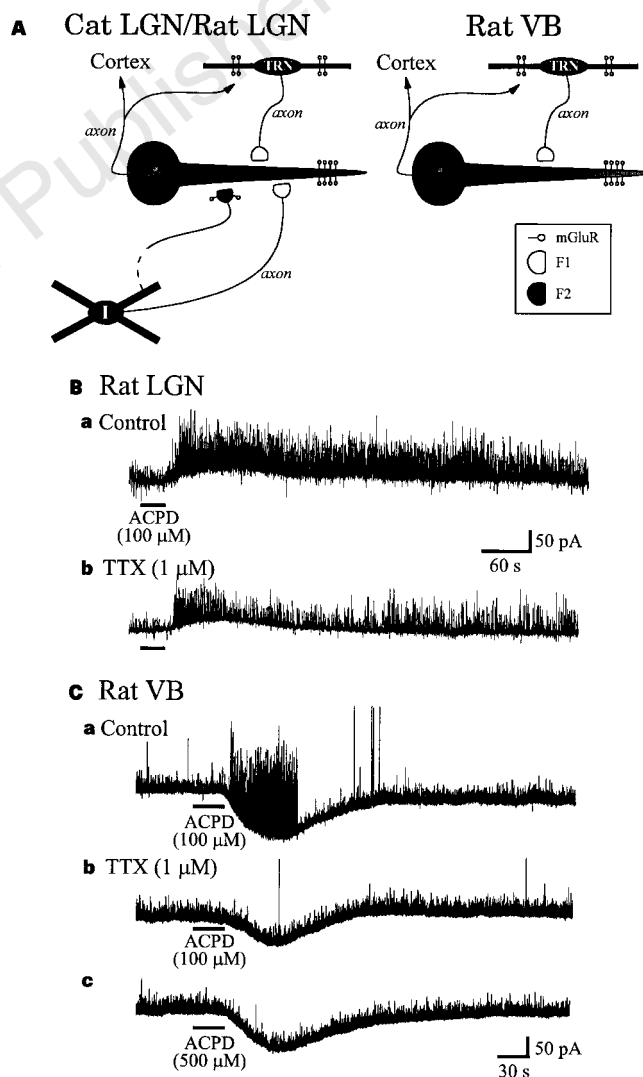


Figure 3 Inhibitory innervation of thalamic relay neurons and voltage clamp recording from rat LGN relay neuron. **A**, Cat LGN/rat LGN: both TRN neurons (TRN) and interneurons (I) contact relay cells (R) via axonal F1 terminals. Dendritic F2 terminals of interneurons also contact relay cells. Rat VB: Rat VB lacks interneurons and F2 terminals, so the relay neuron is inhibited only by F1 terminals. **B**, Voltage clamp recording from rat LGN relay neuron. **a**, ACPD increases sIPSC amplitude and frequency; **b**, in TTX, much of the ACPD-mediated sIPSC increase persists. **C**, Voltage clamp recording from rat VB relay neuron. **a**, ACPD produces a short-lived increase in sIPSCs. **b**, **c**, The ACPD effect is completely blocked in TTX, even with an increased ACPD concentration.

interneurons and F2 terminals^{18,19}. Thus IPSCs recorded in rat VB neurons can only result from activation of F1 axonal terminals. In recordings from relay cells of the rat LGN in control conditions, ACPD produced a robust increase in sIPSC activity (Fig. 3Ba; $n = 19$ cells), and this increase persisted in TTX ($1 \mu\text{M}$) in 5 of 9 cells (Fig. 3Bb). Thus, the TTX-insensitive mGluR effect is present in both rat and cat LGN, nuclei that contain both F1 and F2 terminals.

The effects of ACPD were then tested on rat VB neurons. In control conditions, ACPD increased sIPSC activity in 10 of 17 cells (Fig. 3Ca). However, in contrast to the rat LGN neurons, the ACPD-mediated increase in sIPSC activity was completely suppressed by TTX ($1 \mu\text{M}$) in all nine VB cells tested (Fig. 3Cb, c). These effects of ACPD on rat VB neurons probably represent activation of axons and their F1 terminals from the adjacent TRN. Thus the TTX-insensitive ACPD action was absent in rat VB, a structure containing only F1 terminals, but was present in rat and cat LGNs, each of which contain both F1 and F2 terminals. These results indicate that most or all of the TTX-insensitive increase in sIPSCs we observed in LGN neurons is due to activation of mGluRs on F2 terminals, thereby evoking GABA release and producing inhibitory activity in LGN relay cells.

This conclusion is consistent with immunocytochemical evidence pointing to dense localization of mGluRs on F2 but not F1 terminals in the cat LGN¹². This, in turn, indicates that interneurons can be activated through mGluRs on the dendritic F2 terminals. However, previous recordings made from somata of interneurons revealed no detectable effects of ACPD¹⁵. We confirmed this in somatic recordings from four cat LGN interneurons. Interneuron recordings were identified by their characteristic spike discharge properties¹⁵ and two were further identified morphologically, based on injecting biocytin through the recording electrode. ACPD produced no discernible effect on the interneurons except at very long latencies (1–2 min), when there is an increase in spontaneous excitatory postsynaptic potentials (sEPSPs; Fig. 2Ba). In TTX, there remains only a very long latency, small depolarization (Fig. 2Bb). The cause of these curious long-latency effects of ACPD recorded at the somata will require further investigation, but it seems clear that their latency is too long to be related to the effects attributed to activation of the F2 terminals of these same cells (compare Figs 1–3).

Our findings suggest that the effects of activating F2 terminals synaptically (via mGluRs) are so electronically distant from the cell's soma and axon hillock that they have negligible influence there⁷. The interneuron may thus perform simultaneous and largely independent computation operations: one involves inputs to the proximal dendrites and soma expressed conventionally via the axon; the other involves a collection of local functional circuits via F2 terminals that are both pre- and postsynaptic and do not require action potentials. This indicates that the typical somatic recording of these cells may reveal nothing of the processing involving their F2 terminals. In addition, our results offer insight into the functioning of triadic circuitry. The retinal synapses onto relay cells appear to activate only ionotropic glutamate receptors (iGluRs)¹⁴, whereas, as our data indicate, these retinal terminals can activate mGluRs on the F2 terminals. Whether retinal terminals can also activate iGluRs on F2 terminals is not known. Nonetheless, the differential requirements for activation of iGluRs rather than mGluRs suggest that the relative monosynaptic excitation and disynaptic inhibition evoked by retinal inputs in the relay cells may differ with different firing patterns of the retinal axon. For example, mGluR activation may require higher rates of afferent activity than iGluR activation¹⁴, implying that the disynaptic inhibition develops more strongly as the retinal axons fire at a greater frequency. This would limit the amplitude of retinogeniculate transmission more as retinal cells become more excited. This may prevent geniculate relay cells from saturating in their response to stronger visual stimuli, thereby extending their dynamic range of responsiveness to visual stimuli. □

Methods

Slice preparation. Intracellular recordings were made from relay neurons in cat and rat LGN, and rat VB using whole-cell techniques. Thalamic slices were prepared from young cats (5–8 weeks) and rats (postnatal age, 12–18 days) in compliance with approved animal protocols. Briefly, animals were deeply anaesthetized with ketamine (25 mg kg^{-1} for cats; 33 mg kg^{-1} for rats) and pentobarbital sodium (cats only; 50 mg kg^{-1}). A block of tissue containing the LGN and VB was removed and placed in cold, oxygenated slicing solution containing (in mM): 2.5 KCl, 1.25 NaH_2PO_4 , 10.0 MgCl_2 , 0.5 CaCl_2 , 26 NaHCO_3 , 11.0 glucose and 234.0 sucrose. Thalamic slices ($250\text{--}300 \mu\text{m}$) were cut in a coronal or sagittal plane with a vibrating tissue slicer and placed in a holding chamber (30°C) for >2 h before recording. Individual slices were transferred to a submersion-type recording chamber (cat tissue maintained at 30°C ; rat tissue maintained at room temperature) and continuously perfused with oxygenated physiological solution containing (in mM): 126.0 NaCl, 2.5 KCl, 1.25 NaH_2PO_4 , 2.0 MgCl_2 , 2.0 CaCl_2 , 26.0 NaHCO_3 , 10.0 glucose, pH 7.4. **Whole-cell recording.** Whole-cell recordings were obtained from neurons visualized using a high-power water immersion objective within the slices^{20,21}. The recording pipette solution contained (in mM): 117.0 caesium gluconate, 13.0 CsCl, 1.0 MgCl_2 , 0.07 CaCl_2 , 0.1 ethylene glycol- O,O' -bis(2-aminoethyl)- N,N,N',N' -tetraacetic acid (EGTA), 10.0 N -(2-hydroxyethyl)piperazine- N' -(2-ethanesulphonic acid) (HEPES) and 0.5% biocytin. In some recordings, potassium gluconate and KCl were substituted for Cs gluconate and CsCl. The solutions were adjusted to a final pH of 7.3 and osmolality of 280 mOsm. An Axoclamp 2A amplifier (Axon Instruments) was used in continuous single-electrode voltage clamp mode for current recordings. With Cs^+ -containing recording pipettes, sIPSCs were recorded in voltage clamp mode at a holding potential (V_{hold}) of 0 mV. Hyperpolarizing voltage steps were repetitively given throughout the experiment to monitor any changes in the access resistance (R_{ass}), and recordings were limited to those cells with a stable $R_{\text{ass}} < 20 \text{ M}\Omega$. No correction for liquid junction potential has been made to voltage measurements. ACPD was applied by injecting a bolus into the flow line of the chamber over 30–60 s using a motorized syringe pump. On the basis of the rate of agonist injection and the rate of chamber perfusion, the estimated final bath concentration of ACPD was estimated to be about one-fourth of the concentration introduced in the flow line. All other pharmacological agents were bath-applied.

Received 25 March; accepted 11 May 1998.

- Sherman, S. M. & Guillery, R. W. Functional organization of thalamocortical relays. *J. Neurophysiol.* **76**, 1367–1395 (1996).
- Famiglietti, E. V. Jr & Peters, A. The synaptic glomerulus and the intrinsic neuron in the dorsal lateral geniculate nucleus of the cat. *J. Comp. Neurol.* **144**, 285–334 (1972).
- Guillery, R. W. The organization of synaptic interconnections in the laminae of the dorsal lateral geniculate nucleus of the cat. *Zeitschr. Zellforsch. Mikroskop. Anat.* **96**, 1–38 (1969).
- Hamos, J. E., Van Horn, S. C., Raczkowski, D., Uhlrich, D. J. & Sherman, S. M. Synaptic connectivity of a local circuit neuron in lateral geniculate nucleus of the cat. *Nature* **317**, 618–621 (1985).
- Montero, V. M. Localization of gamma-aminobutyric acid (GABA) in type 3 cells and demonstration of their source to F2 terminals in the cat lateral geniculate nucleus: a Golgi-electron-microscopic GABA-immunocytochemical study. *J. Comp. Neurol.* **254**, 228–245 (1986).
- Ralston, H. J. Evidence for presynaptic dendrites and a proposal for their mechanism of action. *Nature* **230**, 585–587 (1971).
- Bloomfield, S. A. & Sherman, S. M. Dendritic current flow in relay cells and interneurons of the cat's lateral geniculate nucleus. *Proc. Natl Acad. Sci. USA* **86**, 3911–3914 (1989).
- Guillery, R. W. A study of Golgi preparations from the dorsal lateral geniculate nucleus of the adult cat. *J. Comp. Neurol.* **128**, 21–50 (1966).
- Friedlander, M. J., Lin, C.-S., Stanford, L. R. & Sherman, S. M. Morphology of functionally identified neurons in lateral geniculate nucleus of the cat. *J. Neurophysiol.* **46**, 80–129 (1981).
- LeVay, S. & Ferster, D. Relay cell classes in the lateral geniculate nucleus of the cat and the effects of visual deprivation. *J. Comp. Neurol.* **172**, 563–584 (1977).
- Wilson, J. R., Friedlander, M. J. & Sherman, S. M. Fine structural morphology of identified X- and Y-cells in the cat's lateral geniculate nucleus. *Proc. R. Soc. B* **22**, 411–436 (1984).
- Godwin, D. W. *et al.* Ultrastructural localization suggests that retinal and cortical inputs access different metabotropic glutamate receptors in the lateral geniculate nucleus. *J. Neurosci.* **16**, 8181–8192 (1996).
- Chapman, S., Gähwiler, B. H., Do, K.-Q. & Knöpfel, T. Potassium conductances in hippocampal neurons blocked by excitatory amino-acid transmitters. *Nature* **347**, 765–767 (1990).
- McCormick, D. A. & von Krosigk, M. Corticothalamic activation modulates thalamic firing through glutamate "metabotropic" receptors. *Proc. Natl Acad. Sci. USA* **89**, 2774–2778 (1992).
- Pape, H.-C. & McCormick, D. A. Electrophysiological and pharmacological properties of interneurons in the cat dorsal lateral geniculate nucleus. *Neuroscience* **68**, 1105–1125 (1995).
- Lee, K.H. & McCormick, D. A. Modulation of spindle oscillations by acetylcholine, cholecystokinin and $15,15$ -ACP in the ferret lateral geniculate and perigeniculate nuclei *in vitro*. *Neuroscience* **77**, 335–350 (1997).
- Sherman, S. M. & Cox, C. L. Excitatory and inhibitory actions of metabotropic glutamate receptor activation in rat thalamic reticular neurons. *Soc. Neurosci. Abstr.* **23**, 73.10 (1997).
- Arcelli, P., Frasson, C., Regondi, M. C., De Biasi, S. & Spreafico, R. GABAergic neurons in mammalian thalamus: A marker of thalamic complexity? *Brain Res. Bull.* **42**, 27–37 (1997).

19. Ottersen, O. P. & Storm-Mathisen, J. GABA-containing neurons in the thalamus and pretectum of the rodent. An immunocytochemical study. *Anat. Embryol.* **170**, 197–207 (1984).
20. Edwards, F. A., Konnerth, A., Sakmann, B. & Takahashi, T. A thin slice preparation for patch clamp recordings from neurons of the mammalian central nervous system. *Pflügers Arch.* **414**, 600–612 (1989).
21. Stuart, G. J., Dodt, H. U. & Sakmann, B. Patch-clamp recordings from the soma and dendrites of neurons in brain slices using infrared video microscopy. *Eur. J. Physiol.* **423**, 511–518 (1993).

Acknowledgements. This work was supported by National Eye Institute (NIH).

Correspondence and requests for materials should be addressed to S.M.S. (e-mail: ssherman@neurobio.sunysb.edu).

Original antigenic sin impairs cytotoxic T lymphocyte responses to viruses bearing variant epitopes

Paul Klenerman & Rolf M. Zinkernagel

Institute for Experimental Immunology, University Hospital, Schmelzbergstrasse 12, 8091 Zurich, Switzerland

Some viruses, including human immunodeficiency virus (HIV) and hepatitis B virus (HBV) in humans, and lymphocytic choriomeningitis virus (LCMV) in mice, are initially controlled by cytotoxic T lymphocytes (CTLs), but may subsequently escape through mutation of the relevant T-cell epitope^{1–3}. Some of these mutations preserve the normal binding to major histocompatibility complex class I molecules, but present an altered surface to the T-cell antigen receptor^{4,5}. The exact role of these so-called altered peptide ligands *in vivo* is not clear. Here we report that mice primed with LCMV-WE strain respond to a subsequent infection by WE-derived CTL epitope variants with a CTL response directed against the initial epitope rather than against the new variant epitope. This phenomenon of ‘original antigenic sin’ was initially described in influenza^{6–8} and is an asymmetric pattern of protective antibody crossreactivity determined by exposure to previously existing strains, which may therefore extend to some CTL responses. Original antigenic sin by CTL leads to impaired clearance of variant viruses infecting the same individual and so may enhance the immune escape of mutant viruses evolving in an individual host.

We investigated why many CTL-escape viruses arise and persist, despite an ongoing and apparently intact immune response, by using the murine model of LCMV, a non-cytopathic RNA virus infection in which CTLs are responsible both for the initial control of virus replication and for tissue destruction⁹. CTL responses may select epitope variants that escape recognition in LCMV, HIV or HBV infection^{3,10,11}. One such variant virus (LCMV-8.7) arose by selection *in vivo* in C57BL/6 (H-2D^b) mice expressing a transgenic T-cell antigen receptor (TCR)³. This has been shown to bear an altered peptide ligand (APL) at the dominant H-2D^b-restricted CTL epitope in the glycoprotein (GP) (33–41: KAVYNFATC to KALYNFATC; here described as GP 33–3L). C57BL/6 mice infected with wild-type (WT) LCMV-WE (200 plaque-forming units (PFU) intravenously (i.v.) injected) generated CTLs that partially cross-reacted with the mutant epitope from LCMV-8.7, GP33–3L (Fig. 1a, left panels). Mice primed with LCMV-WE (WT), followed by rechallenge with the same LCMV-WE (WT), likewise had an *ex vivo* response directed primarily against the wild-type epitope (Fig. 1a, middle panels). Surprisingly, secondary challenge with LCMV-8.7 also led to CTLs being primarily specific for the priming epitope, rather than for that of the challenge virus (Fig. 1a, right panels). The asymmetric pattern was the same under two other conditions: when restimulation was performed *in vitro* using specific peptides (Fig. 1b, two left panels) and when priming was

done using a recombinant vaccinia virus expressing LCMV-GP (Vac-G2) instead of LCMV itself (Fig. 1b, two right panels). This phenomenon is similar to that described as ‘original antigenic sin’—in which antibody responses stimulated by newly arising mutated but serologically related (drifted) influenza strains are predominantly directed against the first strain encountered^{7,8}—but this has not been observed previously for T-cell responses *in vivo*.

The same phenomenon was analysed for the major CTL epitope NP118, restricted by L^d or L^q, in which a natural variation between LCMV-WE (RPQASGVYM; NP118 WT) and LCMV-Docile (RPQ TSGVYM; NP118-4T) exists, the latter strain having been derived from a neonatally infected LCMV-WE carrier ICR (L^q) mice¹².

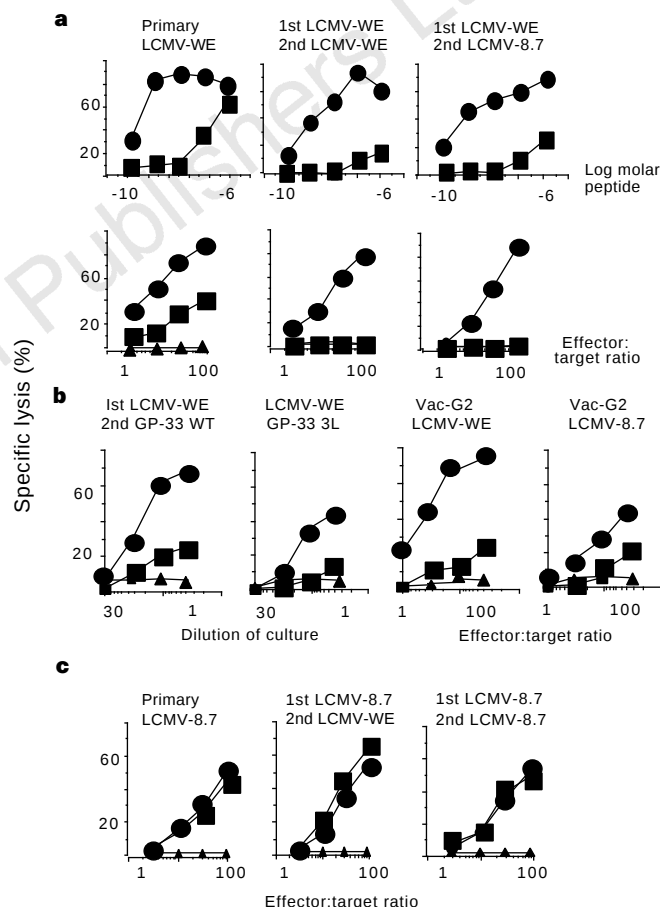


Figure 1 CTL responses of C57BL/6 mice to a single or second virus infection. For all experiments, GP33-WT, GP33-3L and control targets are represented as black circles, squares and triangles respectively. **a**, Primary and secondary responses of C57BL/6 mice to LCMV challenge. Left: mice were infected with LCMV-WE 200 PFU i.v., their splenocytes were collected after 8 days and specific lysis tested against EL4 (H-2D^b) cells sensitized with peptides GP-33 wild type or -3L (upper panels) or against targets prepulsed with peptide at 100 nM (lower panels). Middle and right: C57BL/6 mice were infected with LCMV-WE 200 PFU i.v. and rechallenged with LCMV-WE (middle) or LCMV-8.7 (right). **b**, Responses stimulated with peptide or recombinant vaccinia. Left two panels: splenocytes from C57BL/6 mice previously infected with LCMV-WE were restimulated *in vitro* with GP33-WT (left panel) or 3L variant (right panel) and tested for recognition of peptide-pulsed MC57 targets. Right two panels: the experiment was done as in **a** (middle and right panels) but the initial infection was with Vac-G2 instead of LCMV-WE. **c**, CTL responses of C57BL/6 mice to virus infections in the reverse order (LCMV-8.7 priming). Left: mice were infected with LCMV-8.7 200 PFU i.v., splenocytes were collected after 8 days and specific lysis tested. Middle and right panels: C57BL/6 mice were infected with LCMV-8.7 200 PFU i.v., rechallenged with LCMV-WE or LCMV-8.7 (2×10^6 PFU i.v.) and specific lysis tested.

BALB/c (H-2^d) mice infected 8 days previously with LCMV-WE showed a primary CTL response specific predominantly for NP118-WT (Fig. 2a, left). Again, such mice previously primed with LCMV-WE produced a primarily NP118-WT-specific response, regardless of whether the second challenge was with LCMV-WE or Docile (Fig. 2a, middle and right). Responses were similar for restimulation by 118-WT or -4T peptides *in vitro* (Fig. 2b).

When mice were infected with variant virus, primary CTLs were fully crossreactive on variant and WT epitope-pulsed targets (Fig. 1c, left). When infected in reverse order (variant first and WT second), CTL lysis was equal against both epitopes in all cases (Fig. 1c, middle and right panels). Similar asymmetry was seen for primary responses in BALB/c (H-2L^d) mice and SWR or DBA/1 (H-2L^g) mice (Fig. 2c) and in BALB/c mice primed with LCMV-Docile and

rechallenged with LCMV-WE (data not shown). Thus, in both epitopes the CTL response is determined by the original peptide presented, and crossreactivity is asymmetrical: it is not simply that it is generally difficult to prime naive CTL in the face of established memory responses, instead epitope order determines outcome.

To test whether the effect is the same in mutations arising in mice with a polyclonal CTL repertoire, we used a further mutation, GP33-7T, which was amplified by using the polymerase chain reaction (PCR) and sequenced from an adult C57BL/6 mouse which had been infected with LCMV-WE 60 days previously¹³. This showed the same phenotype as GP33-3L in that it was poorly recognized by CTL from mice previously infected with LCMV-WE, but was nevertheless still able to restimulate these CTL (Fig. 3a). This was not observed for all APL randomly synthesized: of seven non-conservative mutations in GP33 position 4, only one peptide, GP33-4N, could strongly restimulate GP33-specific CTL (data not shown). The natural mutations studied here therefore represent a subgroup of possible APLs with this specific capacity.

The effect was not restricted to immunodominant epitopes, because a second epitope in LCMV-GP in the LCMV-Pasteur strain (for which the exact *in vivo* passage history is not known; SGVENPGGYCL to SGTENPGGYCL: GP276-WT to GP276-3T) is poorly recognized by CTLs from LCMV-WE-infected mice either directly *ex vivo* (data not shown) or after secondary restimulation (Fig. 3b, left). However, as with the other examples, the mutant epitope was able to restimulate effector CTLs *in vitro*, and these

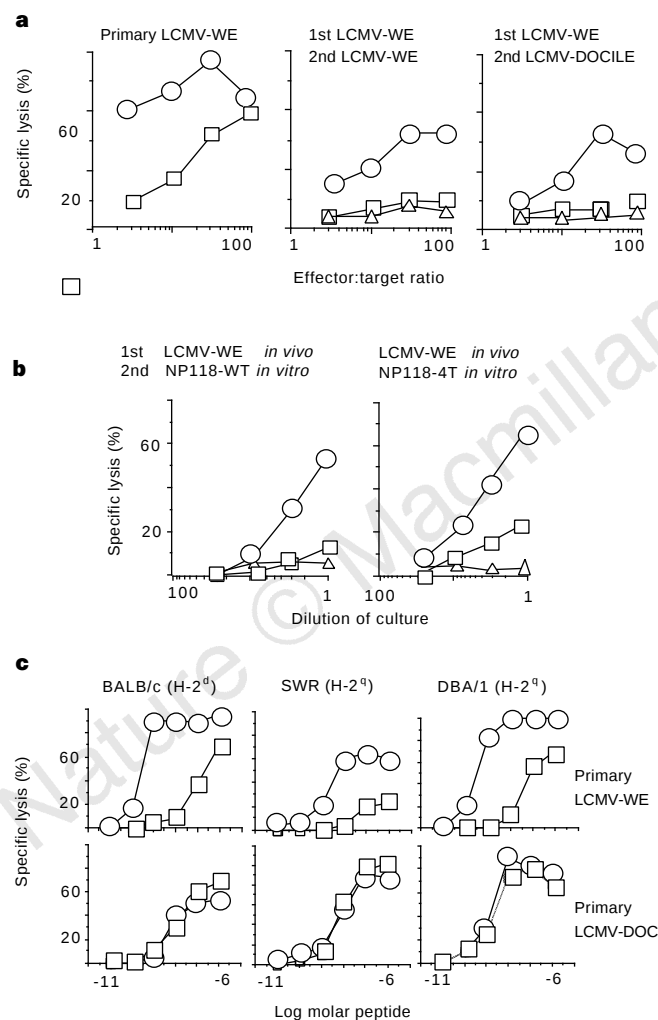


Figure 2 *Ex vivo* responses of BALB/c mice to a single or second virus infection. For all experiments, NP118-WT, NP118-4T and control targets are represented as white circles, squares and triangles respectively. **a**, Primary and secondary responses of BALB/c mice to LCMV challenge. Left, BALB/c mice were infected with LCMV-WE, splenocytes were collected after 8 days and specific lysis was tested against P815 (H-2^d) targets. Middle and right, BALB/c mice were infected with LCMV-WE 200 PFU *i.v.* rechallenged with LCMV-WE (middle) or LCMV-Docile (right) (2×10^6 PFU *i.v.*). **b**, Secondary responses stimulated with peptide *in vitro*. Splenocytes from BALB/c mice previously infected with LCMV-WE were restimulated *in vitro* with NP118 WT (left) or 4T variant (right). **c**, Primary response of BALB/c, SWR and DBA/1 mice to LCMV-WE and Docile challenge. Mice were infected with LCMV-WE (upper panels) or Docile (lower panels) 200 PFU *i.v.*, splenocytes were collected after 8 days, and specific lysis was tested against P815 cells (H-2^d) or DBA/1 cells (H-2^g) sensitized with peptides NP118-WT or -4T.

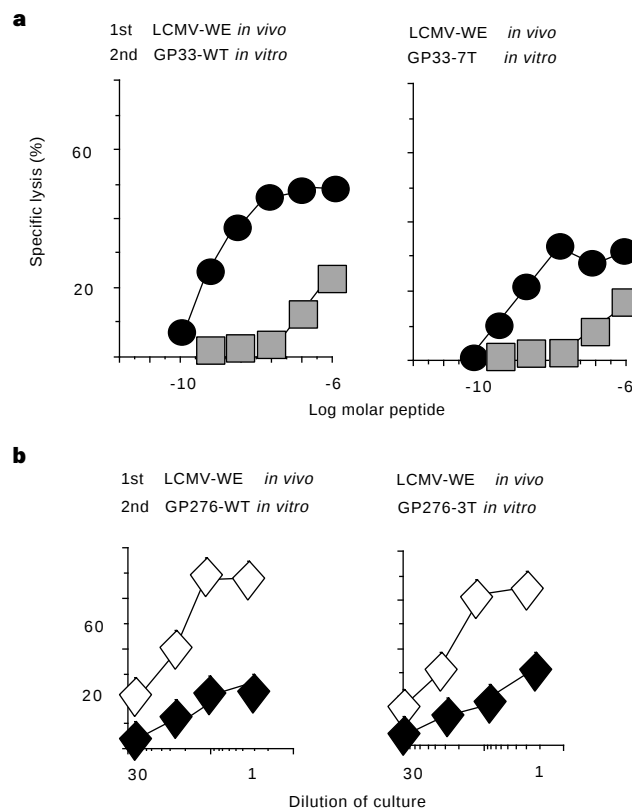


Figure 3 CTL responses to APLs in dominant and subdominant epitopes. **a**, Response of polyclonal CTL to GP33-7T. Splenocytes from mice previously infected with LCMV-WE were restimulated *in vitro* with peptide GP33-WT (left) or -7T (right) and tested against GP33-WT (black circles) or GP33-7T (grey squares) sensitized EL4 cells. **b**, Response to GP276 wild-type and variant epitope GP276-3T. Splenocytes from C57BL/6 mice previously infected with LCMV-WE were restimulated *in vitro* with GP276-WT (left) or GP276-3T peptide (right) and tested against EL4 cells sensitized with peptides GP276-WT (white diamonds) or -3T (black diamonds).

CTLs retained specificity for the wild-type epitope (Fig. 3b, right). In contrast, other CTL epitope mutants with either a non-conservative non-anchor mutation (SGVENPD $\underline{\text{G}}$ GYCL)¹⁴ or anchor mutations leading to loss of binding, were neither recognized by CTLs from LCMV-WE-infected mice, nor were able to restimulate CTLs *in vitro* using identical protocols (data not shown).

We investigated in three ways whether this phenomenon was antigen-dependent or due to bystander effects. First, LCMV-WE-primed C57BL/6 mice made H-2^b-restricted GP- but not the nucleoprotein (NP)-specific CTL responses *ex vivo* 5 days after challenge with a recombinant vaccinia virus bearing LCMV-GP (Vac-G2), and *vice versa* when challenged with a recombinant vaccinia bearing LCMV-NP (Vac-YN4) (Fig. 4a). Second, mice were rechallenged with a mutant LCMV variant (LCMV-WE CL1.2) lacking the second epitope in GP (GP276–86) through mutation of the highly conserved D^b anchor residue asparagine at position 5 (SGVENP $\underline{\text{G}}$ GGYCL to SGVEDP $\underline{\text{G}}$ GGYCL)^{10,11}. Mice primed with LCMV-WE and rechallenged with LCMV-WE CL1.2 failed to make a GP276-specific response, whereas mice rechallenged with LCMV-WE had detectable GP276-specific CTL (Fig. 4b). Third, stimulation *in vitro* with irrelevant D^b and L^d binding peptides derived from adenovirus (SGPSNTPEI) and murine cytomegalovirus (MCMV) (YPHFMPNTL) failed to activate CTLs (data not shown). These results, together with the asymmetry discussed indicate that the biased CTL responses reflect antigen-specific restimulation of the previously primed CTL, and not bystander activation *in vivo* or *in vitro*.

Finally, we investigated whether this biased specificity towards the

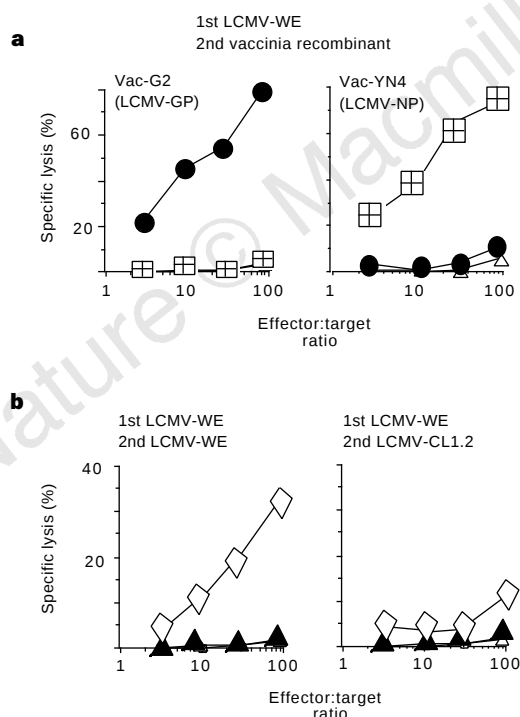


Figure 4 Evaluation of specific versus bystander restimulation *in vivo*. **a**, Specificity of *ex vivo* responses in C57BL/6 mice. C57BL/6 mice were infected with LCMV-WE and then with recombinant vaccinia expressing LCMV-GP (Vac-G2; left) or recombinant vaccinia expressing LCMV-NP (Vac-YN4; right). CTLs were then tested for recognition of EL4 targets prepulsed with D^b-restricted peptides; GP33, NP396 and control targets are represented by circles, crossed squares and triangles, respectively. **b**, Effect of rechallenge with GP276-nonbinding mutant virus. C57BL/6 mice were primed with LCMV-WE 200 PFU i.v. and rechallenged with LCMV-WE or LCMV-WE CL1.2, which does not bind to D^b (refs 10, 11). GP276-pulsed and control targets are represented by white diamonds and black triangles, respectively.

epitope originally encountered could be associated with impaired clearance of variant viruses *in vivo*. This was tested experimentally using LCMV-WE-primed BALB/c mice (H-2^dL^d) which, (like H-2^qL^q mice) produce an anti-LCMV response largely focused on the single NP epitope NP118. These primed mice, after rechallenge *in vivo*, were unable to clear rapidly the variant Docile from liver or spleen after 48 h, although they effectively cleared LCMV-WE (Fig. 5a). By similarly focusing the response in C57BL/6 mice after priming with a recombinant vaccinia virus expressing only GP from WE strain Vac-G2, clearance of LCMV-8.7 (1.1 log PFU per g spleen compared to naive animals) was also reduced compared to LCMV-WE (3.9 log PFU per g spleen compared to naive animals) (Fig. 5b). Finally, priming with LCMV-8.7 led to immediate protection in lung and liver against rechallenge with LCMV-8.7B23 (a mutant LCMV-8.7 selected *in vitro* using H-2^b-restricted CTLs, which also lack the GP276 epitope as a result of mutation of an anchor residue¹¹), whereas priming with LCMV-WE gave initial titres similar to those in naive animals (Fig. 5c).

Our results identify an important phenomenon in the *in vivo* CTL response against variable pathogens and extend the concept of original antigenic sin to CTLs. This was proposed for the directed co-evolution of protective antibodies against influenza virus^{6–8} and

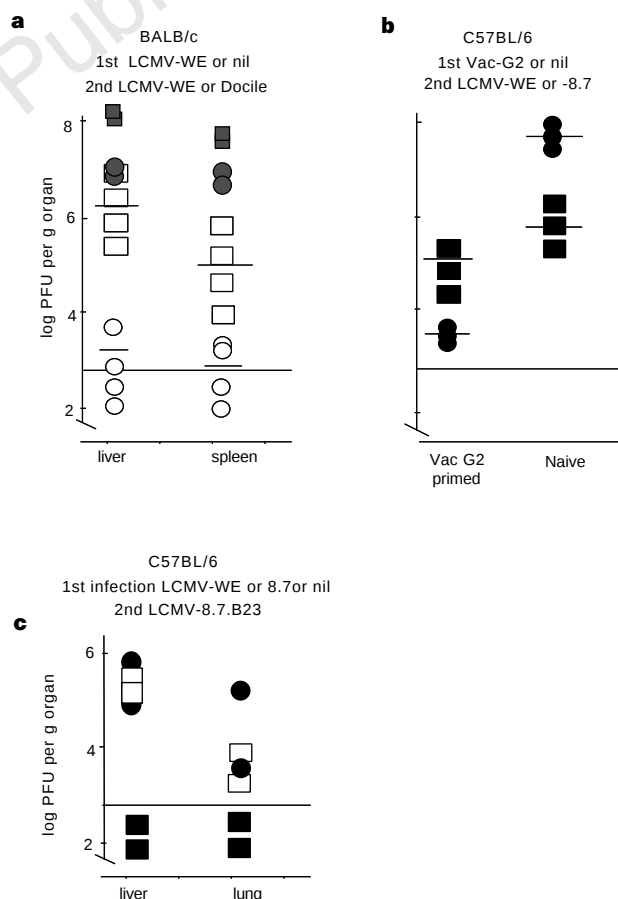


Figure 5 Impaired clearance of LCMV epitope variants from primed mice. **a**, BALB/c mice were primed initially with LCMV-WE 200 PFU i.v. and rechallenged with LCMV-WE (white circles) or Docile (white squares) 2×10^6 PFU i.v. Naive animals were also challenged with LCMV-WE (grey circles) or Docile (grey squares). Titres of virus were determined in organs after 48 h. **b**, C57BL/6 mice were primed initially with Vac-G2 or left uninfected, then rechallenged with LCMV-WE (circles) or -8.7 2×10^6 PFU (squares); splenic viral titres were obtained at 72 h. **c**, C57BL/6 mice were primed initially with LCMV-8.7 (black squares) or LCMV-WE (black circles) or nil (white squares), rechallenged with LCMV-8.7B23 2×10^6 PFU i.v.; titres obtained after 36 h.

subsequently in dengue, HIV and malaria infections^{15–17}. This concept translates readily into current models of presentation of APL to CTLs, leading to segregation of functions through differential signals^{18,19}. Thymocytes receiving such a signal may be positively selected²⁰ and it is possible that encounters of a similar nature may also provoke CTL expansion in the periphery, particularly in a situation, where, as here, primed CTLs are present in excess over naive precursors. A huge expansion of antigen-specific CTLs occurs in LCMV which persist and may be re-expanded during the memory phase^{21–23}. This re-expansion is antigen-specific, but the effect of variant antigens has not previously been explored and such effects may be relevant in emerging HIV, HBV and HCV variants^{24–27}. In the LCMV model used here, this phenomenon was seen in a range of variants arising in mice with monoclonal or polyclonal CTL repertoires, and under a variety of infection conditions. Although adaptability is a dominant feature of the normal immune response, highly mutable viruses may, in a finely tailored way, be testing its limits²⁸. □

Methods

CTLs were prepared from spleens of intravenously infected C57BL/6 and BALB/c mice, and used either directly or after 5 days restimulation *in vitro*, in chromium release assays³. Virus stocks used here LCMV-WE (from F. Lehmann-Grube), LCMV-Docile (from C. Pfau) and LCMV-Armstrong (from M. Buchmeier). For *ex vivo* lysis assays, splenocytes were collected 8 d after acute infection and 4 d after rechallenge (2×10^6 PFU) and used directly in CTL assays²⁹. Challenge with recombinant vaccinia virus was with 2×10^6 PFU i.p. Results are from pooled spleens from 2 mice at a fixed effector:target (E:T) ratio of 50:1 in the case of peptide titrations, and means of 2–3 individual mice in the case of E:T titrations: a representative experiment from 2–3 experiments is shown in each case. Peptides were prepared by solid-phase synthesis (Neosystem, Strasbourg, France). EL4 cells, P815 and MC57G cells were prepulsed at 10–100 nM for use in CTL assays or *in vitro* restimulation. For *in vitro* restimulation, splenocytes from C57BL/6 memory mice stimulated with an irrelevant D^b-restricted peptide derived from adenovirus (SGSPNTPEI), or from BALB/c memory mice stimulated with an irrelevant L^d-restricted peptide derived from MCMV (YPHFMPNTNL), did not generate lytic responses above background level. Virus was titrated by immunological focus assay³⁰ and cDNA was sequenced as before¹³. Recombinant vaccinia viruses Vac-G2 and Vac-YN4 were kindly provided by D. Bishop.

Received 11 February; accepted 28 May 1998.

- Phillips, R. E. *et al.* Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453–459 (1991).
- Bertoletti, A. *et al.* Cytotoxic T lymphocyte response to a wild type hepatitis B virus epitope in patients chronically infected by variant viruses carrying substitutions within the epitope. *J. Exp. Med.* **180**, 933–943 (1994).
- Pircher, H. *et al.* Viral escape by selection of cytotoxic T cell-resistant variants *in vivo*. *Nature* **346**, 629–233 (1990).
- Bertoletti, A. *et al.* Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells. *Nature* **369**, 407–410 (1994).
- Klennerman, P. *et al.* Naturally occurring HIV-1 gag variants antagonize cytotoxic T-cell activity. *Nature* **369**, 403–406 (1994).
- Francis, T. Influenza: The new acquaintance. *Ann. Intern. Med.* **39**, 203–212 (1953).
- De St Groth, F. & Webster, R. Disquisitions on original antigenic sin. I. Evidence in man. *J. Exp. Med.* **124**, 331–345 (1966).
- De St Groth, F. & Webster, R. Disquisitions on original antigenic sin. II. Proof in lower creatures. *J. Exp. Med.* **124**, 347–361 (1966).
- Hotchin, J. The biology of LCMV infection: Virus induced immune disease. *Cold Spring Harbor Symp. Quant. Biol.* **27**, 470–488 (1962).
- Aebischer, T., Moskopidhis, D., Rohrer, U. H., Zinkernagel, R. M. & Hengartner, H. *In vitro* selection of lymphocytic choriomeningitis virus escape mutants by cytotoxic T lymphocytes. *Proc. Natl. Acad. Sci. USA* **88**, 11047–11051 (1991).
- Moskopidhis, D. & Zinkernagel, R. M. Immunobiology of cytotoxic T-cell escape mutants of lymphocytic choriomeningitis virus. *J. Virol.* **69**, 2187–2193 (1995).
- Jacobson, S. & Pfau, C. Viral pathogenesis and resistance to defective interfering particles. *Nature* **283**, 311–315 (1980).
- Klennerman, P., Hengartner, H. & Zinkernagel, R. A non-retroviral RNA virus persists in DNA form. *Nature* **390**, 298–301 (1997).
- Lewicki, H. *et al.* CTL escape variants. I. Generation and molecular characterization. *Virology* **210**, 29–40 (1995).
- Good, M., Zeveing, Y., Currier, J. & Bilsborough, J. Original antigenic sin, T cell memory and malaria sporozoite immunity: a hypothesis for immune evasion. *Parasite Immunol.* **15**, 187–193 (1993).
- Halstead, S., Rojanasuphot, S. & Tsurukubo, Y. Original antigenic sin in dengue. *Am. J. Hyg. Trop. Med.* **32**, 154–156 (1983).
- Nara, P. L. *et al.* Emergence of viruses resistant to neutralization by V3-specific antibodies in experimental human immunodeficiency virus type 1 IIIB infection of chimpanzees. *J. Virol.* **64**, 3779–

- 3791 (1990).
- Evavold, B. & Allen, P. Separation of IL-4 production from proliferation by an altered peptide ligand. *Science* **252**, 1308–1310 (1991).
- Kessler, B., Bassanini, P., Cerottini, J. & Luescher, I. Effects of epitope modification on TCR ligand binding and antigen recognition by seven H-2K^d restricted CTL clones specific for a photoreactive peptide derivative. *J. Exp. Med.* **185**, 629–640 (1997).
- Jameson, S. C. & Bevan, M. J. T cell receptor antagonists and partial agonists. *Immunity* **2**, 1–11 (1995).
- Butz, E. & Bevan, M. Massive expansion of antigen-specific CD8⁺ cells during acute virus infection. *Immunity* **8**, 167–175 (1998).
- Murali-Krishna, K. *et al.* Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity* **8**, 177–187 (1998).
- Gallimore, A. *et al.* Induction and exhaustion of LCMV-specific CTL visualised using soluble tetrameric MHC–peptide complexes. *J. Exp. Med.* **187**, 1383–1393 (1998).
- Klennerman, P., Meier, U. C., Phillips, R. E. & McMichael, A. J. The effects of natural altered peptide ligands on the whole blood cytotoxic T lymphocyte response to human immunodeficiency virus. *Eur. J. Immunol.* **25**, 1927–1931 (1995).
- McAdam, S. N. *et al.* Immunogenic HIV variants that bind to HLA-B8 but fail to stimulate CTL responses. *J. Immunol.* **155**, 2729–2736 (1995).
- Weiner, D. *et al.* Persistent hepatitis C virus infection in a chimpanzee is associated with emergence of a CTL escape variant. *Proc. Natl. Acad. Sci. USA* **82**, 2755–2759 (1995).
- Kent, S., Greenberg, P., Hoffman, M., Akridge, R. & McElrath, M. Antagonism of vaccine induced HIV-1 specific CD4 T cells by primary HIV-1 infection. Potential mechanism of vaccine failure. *J. Immunol.* **158**, 807–815 (1997).
- Zinkernagel, R. M. Immunology taught by viruses. *Science* **271**, 173–178 (1996).
- Bachmann, M. F., Speiser, D. & Ohashi, P. Functional maturation of an antiviral CTL response. *J. Virol.* **71**, 5764–5768 (1997).
- Battegay, M. *et al.* Quantification of LCMV with immunological focus assay in 24 or 96 well plates. *J. Virol. Meth.* **33**, 191–198 (1991).

Acknowledgements. This work was sponsored by the Wellcome Trust, the Swiss National Foundation for Science and the Kanton of Zurich. We thank E. Horvath for technical work, and P. Aichele, M. Bachmann, S. Ehl and A. McPherson for discussion.

Correspondence and requests for materials should be addressed to R.M.Z. (e-mail:rolf.zinkernagel@pty.unizh.ch).

Role of HIF-1 α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis

Peter Carmeliet*, Yuval Dor†, Jean-Marc Herbert‡, Dai Fukumura§, Koen Brusselmans*, Mieke Dewerchin*, Michal Neeman||, Françoise Bono‡, Rinat Abramovitch||, Patrick Maxwell†, Cameron J. Koch#, Peter Ratcliffe†, Lieve Moons*, Rakesh K. Jain§, Désiré Collen* & Eli Keshet†

* Center for Transgene Technology and Gene Therapy, Flanders Interuniversity Institute for Biotechnology, KU Leuven, Leuven B-3000, Belgium

† Department of Molecular Biology, Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel

‡ Sanofi Recherche, Haemobiology Research Department, Toulouse Cedex 31036, France

§ Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

|| Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel

¶ Institute of Molecular Medicine, John Radcliffe Hospital, Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK

Department of Radiation Oncology, School of Medicine, University of Pennsylvania, Philadelphia 19104-6003, USA

As a result of deprivation of oxygen (hypoxia) and nutrients, the growth and viability of cells is reduced¹. Hypoxia-inducible factor (HIF)-1 α helps to restore oxygen homeostasis by inducing glycolysis, erythropoiesis and angiogenesis^{2–4}. Here we show that hypoxia and hypoglycaemia reduce proliferation and increase apoptosis in wild-type (HIF-1 α ^{+/+}) embryonic stem (ES) cells, but not in ES cells with inactivated HIF-1 α genes (HIF-1 α ^{-/-}); however, a deficiency of HIF-1 α does not affect apoptosis induced by cytokines. We find that hypoxia/hypoglycaemia-regulated genes involved in controlling the cell cycle are either HIF-1 α -dependent (those encoding the proteins p53, p21, Bcl-2) or HIF-1 α -independent (p27, GADD153), suggesting that there are at

least two different adaptive responses to being deprived of oxygen and nutrients. Loss of HIF-1 α reduces hypoxia-induced expression of vascular endothelial growth factor, prevents formation of large vessels in ES-derived tumours, and impairs vascular function, resulting in hypoxic microenvironments within the tumour mass. However, growth of HIF-1 α tumours was not retarded but was accelerated, owing to decreased hypoxia-induced apoptosis

and increased stress-induced proliferation. As hypoxic stress contributes to many (patho)biological disorders^{1,5}, this new role for HIF-1 α in hypoxic control of cell growth and death may be of general pathophysiological importance.

HIF-1 α ^{-/-} ES cells were generated by homologous recombination and subsequent selection in high G418 (Fig. 1a, b). ES clones harbouring a randomly integrated HIF-1 α gene targeting vector

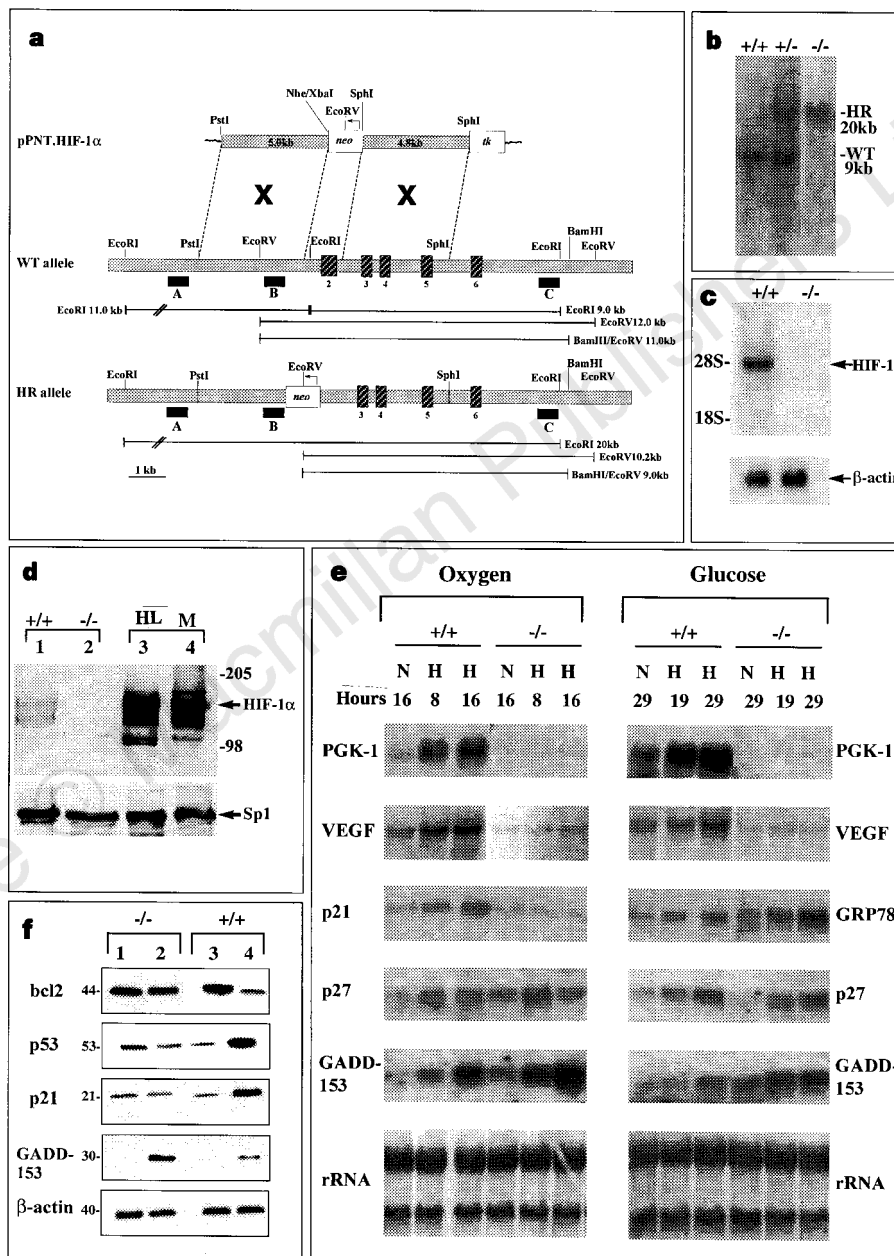


Figure 1 Targeting of the HIF-1 α gene. **a**, Strategy to disrupt the HIF-1 α gene. Top, targeting vector pPNT.HIF-1 α ; middle, map of the wild-type (WT) HIF-1 α gene; bottom, homologously recombined (HR) HIF-1 α allele with numbered exons (dark blocks) and introns (grey shaded bars). Hybridization probes A (0.5-kb *Hinc2*/PstI fragment), B (1.5-kb *StuI*/NheI fragment) and C (0.4-kb *Afl* fragment) and analytical restriction digests are indicated. **b**, Southern blot of *Eco*RI-digested genomic DNA from ES cell clones (probe C) generating a 9.0-kb WT and a 20.0-kb HR HIF-1 α allele. **c**, Northern blot analysis of total mRNA from ES cells hybridized with a murine HIF-1 α exon 2-specific cDNA probe or a β -actin probe. **d**, Western blot analysis for HIF-1 α or the ubiquitous transcription factor Sp1 (control) on nuclear extracts from rHIF-1 α ^{+/+} ES cells (lane 1), HIF-1 α ^{-/-} ES cells (lane 2), human HeLa (HL) cells (lane 3) or mouse Hepa1 cells (M) (lane 4), treated with

100 μ M desferrioxamine (DFO; 16 h). **e**, Northern blot analysis of total mRNA derived from undifferentiated rHIF-1 α ^{+/+} (+/+) or HIF-1 α ^{-/-} (-/-) ES cells during normoxia (N; 20% O₂) or hypoxia (H; 2% O₂), or during normoglycaemia (N; 25 mM glucose) or hypoglycaemia (H; 0 mM glucose) for the indicated time (hours). PGK-1, phosphoglycerokinase; VEGF: vascular endothelial growth factor; GADD153: growth arrest and DNA-damage-inducible gene 153; GRP78: glucose-regulated protein 78. The rRNA is shown as a control of loading. **f**, Western blot analysis on total cell extract from HIF-1 α ^{-/-} (lanes 1, 2) or rHIF-1 α ^{+/+} ES cells (lanes 3, 4) during normoxia (lanes 1, 3), during anoxia/hypoglycaemia (4 h; followed by recovery for 20 h in control conditions) for Bcl-2, p53, p21 (lanes 2, 4), or during hypoxia (2% O₂; 24 h) for GADD153 (lanes 2, 4); β -actin was used as control.

(rHIF-1 $\alpha^{+/+}$ cells) which survived high G418 selection were used as controls. Loss of HIF-1 α gene function was confirmed by northern blotting (Fig. 1c), western blotting (Fig. 1d), and electrophoretic mobility shift assay (EMSA) (see also Fig. S1 in Supplementary Information, which is denoted by the prefix 'S'). Expression of known HIF-1 α target genes (*VEGF*, *PGK-1*, *LDH-A*, *GLUT-1*)³ was induced by hypoxia or hypoglycaemia in rHIF-1 $\alpha^{+/+}$ ES cells but barely occurred in HIF-1 $\alpha^{-/-}$ ES cells (Figs 1e and S2). Both HIF-1 α -dependent (*PGK-1* (ref. 3), *VEGF* (ref. 3), *p21* (ref. 6)) and HIF-1 α -independent (*p27* (ref. 7), *GADD153* (ref. 8), *GRP78* (ref. 9)) regulation of target genes was observed, indicating that there must be at least two pathways for regulating hypoxic genes.

We studied hypoxia/hypoglycaemia-driven apoptosis by counting oligonucleosomes and cells labelled with terminal deoxynucleotidyl transferase-mediated deoxyuridine nick end-labelling (TUNEL). Apoptosis was comparable in rHIF-1 $\alpha^{+/+}$ and HIF-1 $\alpha^{-/-}$ ES cells during normoxia/normoglycaemia and increased to a similar extent after stimulation with a cytokine mixture (Table 1). In contrast, during normoglycaemia/hypoxia, hypoglycaemia/normoxia, normoxia in the presence of 2-deoxyglucose, or hypoglycaemia/anoxia, apoptosis increased 10–30-fold in rHIF-1 $\alpha^{+/+}$ ES cells but was unaffected in HIF-1 $\alpha^{-/-}$ ES cells (Table 1). Hypoxia also induced apoptosis in wild-type C4.5 cells¹⁰ but not in HIF-1 α -deficient Ka13 Chinese hamster ovary (CHO) cells¹⁰ (Table S1), indicating that

HIF-1 α could be a general mediator of hypoxia/hypoglycaemia-driven apoptosis. The levels of p53 (a mediator of a genotoxic apoptosis^{11,12} that is upregulated during hypoxia¹³) and of p21 (a p53 target gene and effector of cell-cycle arrest⁶) were significantly increased, whereas the amount of the apoptosis-inhibitor Bcl-2 (ref. 14) was reduced in stressed rHIF-1 $\alpha^{+/+}$ but not in HIF-1 $\alpha^{-/-}$ ES cells (Fig. 1e, f and Table 1). Conversely, hypoxia-induced levels of the pro-apoptotic and growth-arrest gene product GADD153 (ref. 8) and of the cyclin-dependent kinase inhibitor p27 (ref. 7) occurred independently of HIF-1 α (Fig. 1e, f).

Hypoxic stress suppressed proliferation of rHIF-1 $\alpha^{+/+}$ ES cells, cultured either as undifferentiated monolayers or as differentiated three-dimensional embryoid bodies (Table 1). In contrast, it did not alter proliferation of HIF-1 $\alpha^{-/-}$ ES cell monolayers, but significantly increased proliferation in HIF-1 $\alpha^{-/-}$ embryoid bodies (Table 1). Thus, HIF-1 α deficiency prevents cells from undergoing hypoxic growth arrest and, in the appropriate (mitogenic) microenvironment, stimulates proliferation. In previous studies, loss of HIF-1 α ^{4,15} or of its heterodimer HIF-1 β /ARNT¹⁶ did not significantly alter or reduce the accumulation of cells during hypoxia or hypoglycaemia, although proliferation and apoptosis were not measured directly. Whether, and to what extent, the role of HIF-1 α and of HIF-1 β in cell proliferation and apoptosis depends on genotype, cell type or experimental conditions such as pH¹⁷, cell density,

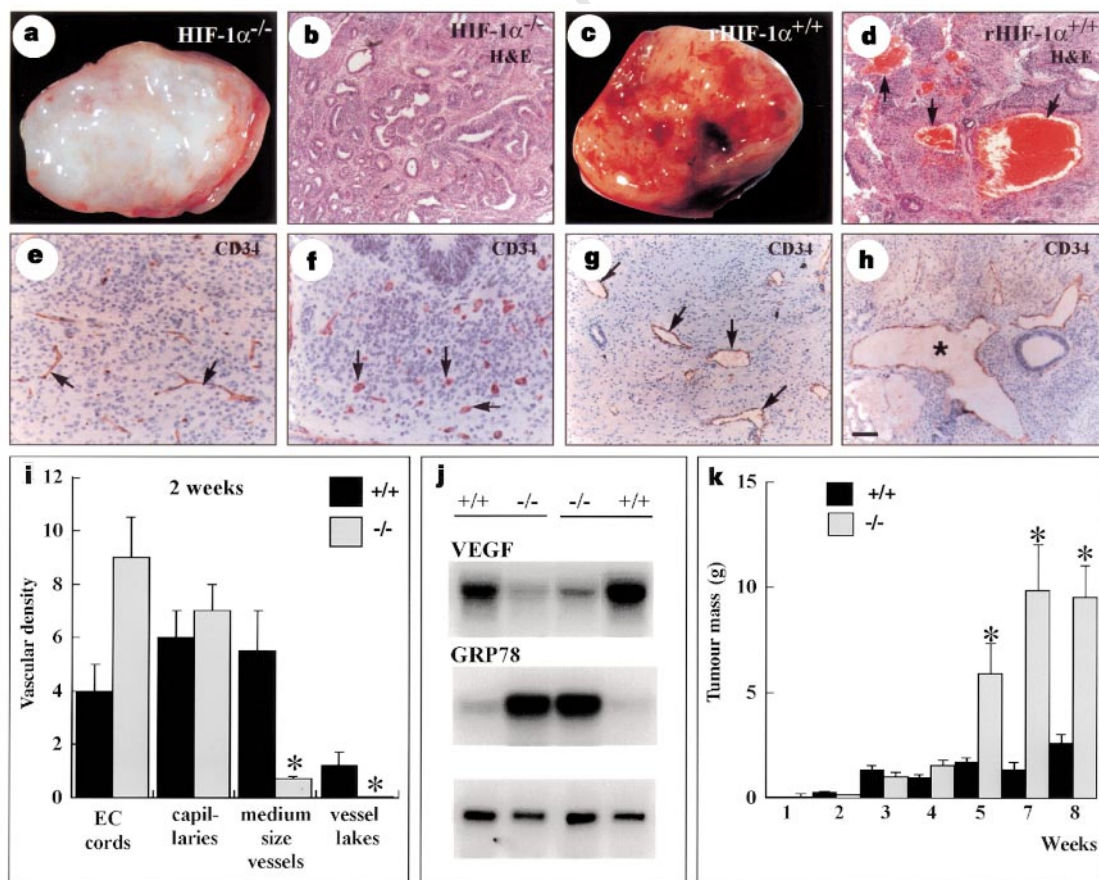


Figure 2 Vascularization of HIF-1 $\alpha^{-/-}$ and rHIF-1 $\alpha^{+/+}$ tumours. **a–d**, Macroscopic (**a, c**) and haematoxylin–eosin stained (**b, d**) sections through HIF-1 $\alpha^{-/-}$ (**a, b**) and rHIF-1 $\alpha^{+/+}$ (**c, d**) tumours revealing the red/haemorrhagic phenotype in rHIF-1 $\alpha^{+/+}$ tumours which contrast to the white and poorly vascularized HIF-1 $\alpha^{-/-}$ tumours. Note the large blood vessels in **d** (arrows) and their absence in **b**. The blood around the HIF-1 $\alpha^{-/-}$ tumour (**a**) has leaked from blood vessels in its surrounding capsule. **e–h**, CD34-immunostained sections through HIF-1 $\alpha^{-/-}$ (**e, f**) and rHIF-1 $\alpha^{+/+}$ (**g, h**) tumours revealing the presence of endothelial cords (<2 μ m; arrows in **e**) and capillaries (~8 μ m; arrows in **f**) in HIF-1 $\alpha^{-/-}$ tumours, and the presence of

medium-sized vessels (~30 $\times$ 70 μ m; arrows in **g**) and large blood vessel lakes ~90 $\times$ 280 μ m; asterisk in **h**) in the rHIF-1 $\alpha^{+/+}$ tumours. **i**, Vascular density (mean $\pm$ s.e.m.) of the different types of blood vessels per optical field in 7 or 8 two-week-old tumours. Asterisks, $P < 0.05$ versus rHIF-1 $\alpha^{+/+}$ by unpaired Student's t -test. Similar data were obtained at 4 and 7 weeks (Table S2). **j**, Northern blot analysis from rHIF-1 $\alpha^{+/+}$ and HIF-1 $\alpha^{-/-}$ tumours for VEGF, GRP78 and β -actin (control). **k**, Growth rate of HIF-1 $\alpha^{-/-}$ (shaded bars) and rHIF-1 $\alpha^{+/+}$ (black bars) tumours. Asterisks, $P < 0.05$ versus rHIF-1 $\alpha^{+/+}$. The bar in **h** represents 50 μ m in **e, f**; 100 μ m in **g, h**; 200 μ m in **b, d**; and 25 mm in **a, c**.

Table 1 Apoptosis and proliferation in HIF-1 α ^{-/-} and rHIF-1 α ^{+/+} ES cells

			Normoxia/ normoglycaemia	Hypoxia/ normoglycaemia	Normoxia/ hypoglycaemia	Normoxia/ 2-deoxyglucose	Anoxia/ hypoglycaemia
Apoptosis	Oligonucleosomes	rHIF-1 α ^{+/+}	32 ± 5	370 ± 20*	130 ± 11*	130 ± 12*	1,300 ± 120*
		+ cytokines	360 ± 12*	ND	ND	ND	ND
		HIF-1 α ^{-/-}	34 ± 3	31 ± 2	41 ± 1	37 ± 4	45 ± 6
	TUNEL-positive cells	+ cytokines	380 ± 20*	ND	ND	ND	ND
		rHIF-1 α ^{+/+}	4.3 ± 0.19	36 ± 2*	26 ± 1.8*	23 ± 2*	61 ± 4*
		HIF-1 α ^{-/-}	3.7 ± 0.12	4.5 ± 0.2	7.2 ± 1.7	5 ± 1	7 ± 2
	Bcl-2	rHIF-1 α ^{+/+}	100 ± 12	27 ± 4*	31 ± 4*	45 ± 1.9*	11 ± 1*
		HIF-1 α ^{-/-}	110 ± 9	120 ± 12	96 ± 9	120 ± 7.3	110 ± 16
	p21	rHIF-1 α ^{+/+}	0.41 ± 0.1	1.2 ± 0.2*	0.93 ± 0.1*	0.74 ± 0.2*	1.5 ± 0.2*
		HIF-1 α ^{-/-}	0.34 ± 0.1	0.33 ± 0.1	0.32 ± 0.1	0.37 ± 0.1	0.3 ± 0.1
Proliferation	p53	rHIF-1 α ^{+/+}	0.11 ± 0.03	3.1 ± 0.1*	2.4 ± 0.1*	2.2 ± 0.04*	3.9 ± 0.13*
		HIF-1 α ^{-/-}	0.12 ± 0.04	0.13 ± 0.04	0.11 ± 0.07	0.10 ± 0.03	0.14 ± 0.13
	c.p.m. ³ H-thymidine	rHIF-1 α ^{+/+}	3,500 ± 400	2,500 ± 300*	ND	ND	ND
		HIF-1 α ^{-/-}	3,600 ± 470	3,400 ± 380	ND	ND	ND
	BrdU-positive cell fraction	rHIF-1 α ^{+/+}	24 ± 4	5 ± 1*	ND	ND	9 ± 2*
		HIF-1 α ^{-/-}	23 ± 5	46 ± 8*	ND	ND	50 ± 10*

The data represent the mean ± s.d. (*n* is 9 to 12) of the number of oligonucleosomes, the units of Bcl-2 and p21, or the pg amounts of p53 antigen, all per 10⁵ cells; or the mean ± s.e.m. (*n*, 8–10) of the fraction of TUNEL- or BrdU-positive cells (per cent of total), or the amount of c.p.m. ³H-thymidine incorporated per 24 well. ND, not done.

* Statistical significance (*P* < 0.05) versus control (normoxia/normoglycaemia without cytokines) by unpaired Student's *t*-test.

oxygen or glucose concentrations, or the amount of serum or growth factors (unpublished observations), some of which regulate apoptosis induced by other types of stress^{12,18,19} is still unclear.

To extend our *in vitro* data, rHIF-1 α ^{+/+} and HIF-1 α ^{-/-} ES cells were subcutaneously injected into nude mice. HIF-1 α ^{-/-} tumours (44 analysed) were most often (75%) homogeneously white, with minimal bleeding (Fig. 2a), and less frequently (25%) slightly haemorrhagic. In contrast, rHIF-1 α ^{+/+} tumours (33 analysed) usually (82%) bled profusely and severely, with fewer tumours (18%) being mildly haemorrhagic (Fig. 2c). HIF-1 α ^{-/-} tumours lacked medium-sized and large vessels but not endothelial cords or capillaries (Fig. 2b, d, e–i; Table S2). Magnetic resonance imaging (MRI) and intravital microscopy (Fig. 3a–d) revealed that HIF-1 α ^{-/-} tumours had a uniform vascularization pattern, with fewer large blood vessels and more avascular zones than rHIF-1 α ^{+/+} tumours, which were characterized by a variable and irregular vascular network, typical of many tumours²⁰. The intravital haemodynamics in rHIF-1 α ^{+/+} tumours compared with HIF-1 α ^{-/-} tumours were (mean ± s.e.m.): vessel density (total length of vessels per unit area in cm cm⁻²), 120 ± 16 (*n* = 18) versus 86 ± 14 (*n* = 25; *P* < 0.05 by ANOVA); vessel diameter (μm), 21 ± 2 (*n* = 16) versus 14 ± 1 (*n* = 21; *P* < 0.05); and flow rate (picolitres s⁻¹), 220 ± 50 (*n* = 44) versus 90 ± 13 (*n* = 44; *P* < 0.05). This vascular phenotype may be due to a lack of upregulation of the expression of vascular endothelial growth factor (VEGF) in hypoxic areas of HIF-1 α ^{-/-} tumours (Figs 2j and S3a, c, g)²¹.

Injection of microspheres revealed a lower blood flow through HIF-1 α ^{-/-} than through rHIF-1 α ^{+/+} tumours (0.079 ± 0.0027 ml min⁻¹ g⁻¹ versus 0.16 ± 0.046 ml min⁻¹ g⁻¹; *n* = 6; *P* = 0.013 by paired Student's *t*-test). MRI analysis was used to measure blood flow and changes in oxygenated haemoglobin in response to changes in inspired oxygen. Perfusion and oxygenation were lower in HIF-1 α ^{-/-} than in rHIF-1 α ^{+/+} tumours, as evidenced by the 10-fold lower *b*ΔY values (where *b* is ml blood per ml tissue and ΔY is the change in oxyhaemoglobin fraction/total haemoglobin; see Methods) (Fig. 3a, b): the mean enhancement of *b*ΔY (a parameter of general O₂ delivery) was 0.014 ± 0.002 in rHIF-1 α ^{+/+} tumours versus 0.0016 ± 0.0008 in HIF-1 α ^{-/-} tumours (*n* = 4; *P* < 0.001 by *t*-test; Table S3). Optical measurement of the pO₂ revealed consistently lower values in HIF-1 α ^{-/-} than in rHIF-1 α ^{+/+} tumours (Fig. 3e). HIF-1 α ^{-/-} tumours also had 3-fold more hypoxic areas than did rHIF-1 α ^{+/+} tumours (Fig. 3f, g), as revealed by

staining for the hypoxia marker EF5 (fractional EF5-positive area: 3.6 ± 0.7% versus 1.1 ± 0.4%, respectively; *n* = 11; *P* < 0.005). Whereas levels of the HIF-1 α target gene encoding VEGF (frequently co-expressed with GRP78 in microenvironments close to necrosis; Fig. 2j and Fig. S3a–d) were higher in rHIF-1 α ^{+/+} than in HIF-1 α ^{-/-} tumours, expression of the HIF-1 α -independent stress-induced marker GRP78 (and GADD153; not shown) was stronger and more widespread in HIF-1 α ^{-/-} than in rHIF-1 α ^{+/+} tumours (Figs 2j and S3b, g, h), indicating that hypoxic/hypoglycaemic stress was greater in HIF-1 α ^{-/-} than in rHIF-1 α ^{+/+} tumours.

Although the diminished vascular supply would be expected to reduce the rate of HIF-1 α ^{-/-} tumour growth, growth rates were comparable during the first 3 weeks, being similar to what has been observed for ARNT^{-/-} ES cell tumour growth (C. Simon, personal communication), but then growth of HIF-1 α ^{-/-} tumours far exceeded that of rHIF-1 α ^{+/+} tumours (Fig. 2k). Cellular proliferation (measured as the percentage of PCNA-positive cells, where PCNA is proliferating cell nuclear antigen) was comparable in 2-week-old tumours (4.7 ± 2.3% in rHIF-1 α ^{+/+} versus 8.1 ± 2.3% in HIF-1 α ^{-/-} tumours; *P* = NS), but at 4 to 7 weeks was significantly higher in HIF-1 α ^{-/-} tumours (11 ± 2% and 18 ± 2%, respectively) than in rHIF-1 α ^{+/+} tumours (3.4 ± 1.6% and 3.4 ± 1.0%; *P* < 0.05). In rHIF-1 α ^{+/+} tumours, cell proliferation in 134 EF5-positive hypoxic fields was absent (in 49%) or minimal (in 51%) (as indicated by the few PCNA-positive cells in a stressed GRP78-positive rHIF-1 α ^{+/+} field; Fig. S3e). In contrast, in HIF-1 α ^{-/-} tumours, hypoxia appeared directly to stimulate proliferation, as 75% of 157 EF5-positive fields were intensely labelled for PCNA (as shown by the many PCNA-labelled cells within a stressed GRP78-positive HIF-1 α ^{-/-} field in Fig. S3i). Thus, similar genotypic differences in hypoxic control of proliferation were observed in tumours and in embryoid bodies. In addition, loss of HIF-1 α affected tumour growth by progressive selection of intensely proliferating cells, as occurs in the absence of p53 (ref. 22). Indeed, PCNA-labelled cells (predominantly epitheloid cells) accumulated more in HIF-1 α ^{-/-} tumours than in rHIF-1 α ^{+/+} tumours (fractional epitheloid area: 26 ± 8%, 46 ± 4%, 46 ± 3% in HIF-1 α ^{-/-} tumours, and 37 ± 6%, 18 ± 3%, 9 ± 2% in rHIF-1 α ^{+/+} tumours at 2, 4 and 7 weeks, respectively; *P* < 0.05 versus HIF-1 α ^{-/-} at 4 and at 7 weeks; Fig. S3m, n).

Apoptosis (quantified by the number of oligonucleosomes per mg protein) was consistently lower in HIF-1 α ^{-/-} tumours

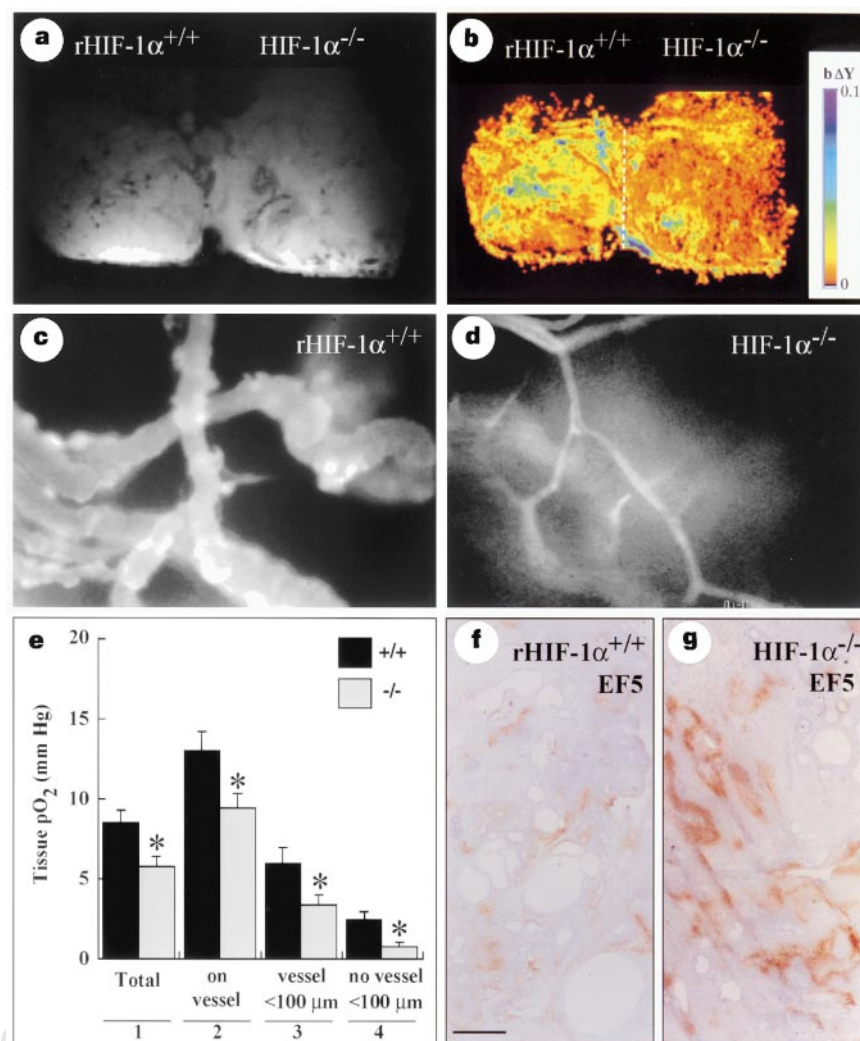


Figure 3 Functional vascularization and growth of HIF-1 $\alpha^{-/-}$ and rHIF-1 $\alpha^{+/+}$ tumours. **a, b**, Gradient echo magnetic resonance imaging (MRI) of a rHIF-1 $\alpha^{+/+}$ (left) and HIF-1 $\alpha^{-/-}$ (right) tumour during inhalation of 95% oxygen and 5% CO₂. **a**, Greyscale original MRI image; **b**, calculated 'functional vascular' ($b\Delta Y$) map of the same tumours. Colour code of $b\Delta Y$: minimum '0' (red) indicates undetectable oxygen delivery; maximum '0.1' (purple) reflects the degree of oxygen delivery in a large perfused vessel. For details, see Methods. The HIF-1 $\alpha^{-/-}$ and rHIF-1 $\alpha^{+/+}$ tumours were mounted so that their overlying skin appears merged on the image (white dotted line). **c, d**, FITC fluorescence images of microvessels in a rHIF-1 $\alpha^{+/+}$ (**c**) and HIF-1 $\alpha^{-/-}$ (**d**) tumour by intravital fluorescence microscopy after

intravenous injection of FITC-dextran. **e**, Tissue pO_2 values (mean $\pm$ s.e.m.) in rHIF-1 $\alpha^{+/+}$ and HIF-1 $\alpha^{-/-}$ tumours. (1) 'Total': average pO_2 throughout the entire tumour; (2) 'on vessel': blood vessel within high-resolution tissue pO_2 sampling field (40 μ m in diameter); (3) 'vessel < 100 μ m': blood vessels within 100 μ m from tissue pO_2 sampling field; (4) 'no vessel < 100 μ m': no blood vessel within 100 μ m from tissue pO_2 sampling field. Asterisks, $P < 0.05$ versus rHIF-1 $\alpha^{+/+}$ by unpaired t -test. **f, g**, EF5 staining of a rHIF-1 $\alpha^{+/+}$ (**f**) and HIF-1 $\alpha^{-/-}$ (**g**) tumour, revealing more, and more strongly, stained hypoxic areas in the HIF-1 $\alpha^{-/-}$ tumour. Scale bar, 200 μ m in **f, g**; 100 pixels in **c, d** represent 100 μ m; tumours in **a, b** weighed ~ 3 g each.

(870 $\pm$ 50, 1,600 $\pm$ 100 and 2,500 $\pm$ 150 at 2, 4 and 7 weeks) than in rHIF-1 $\alpha^{+/+}$ tumours (1,500 $\pm$ 130, 2,200 $\pm$ 200 and 3,100 $\pm$ 100 at 2, 4 and 7 weeks; $P < 0.05$). Most (85%) TUNEL-positive fields were also EF5-positive in tumours of both genotypes (Fig. S3k, l), confirming that apoptosis *in vivo* is hypoxia-driven. Whereas tumours of both genotypes had comparable TUNEL/EF5-positive fields (74% of 134 EF5-positive fields in rHIF-1 $\alpha^{+/+}$ tumours, 80% of 160 EF5-positive zones in HIF-1 $\alpha^{-/-}$ tumours), fewer apoptotic cells were present in HIF-1 $\alpha^{-/-}$ tumour fields (2.5 $\pm$ 0.2% versus 1.5 $\pm$ 0.2% cross-sectional area; $P < 0.05$; Fig. S3f, j). The less obvious hypoxic apoptosis *in vivo* than *in vitro* may relate to differences in environmental signals, influencing the switch towards growth arrest and survival rather than apoptosis, as shown for apoptosis induced by other types of stress^{12,18}.

Many tumours contain hypoxic microenvironments, a condition that is associated with poor prognosis and resistance to

treatment^{20,23}. Our findings indicate that tumour vascularization is largely controlled by HIF-1 α , in part as a result of upregulation of VEGF. In addition, a new role for HIF-1 α in hypoxic control of growth and apoptosis has been revealed, the *in vivo* significance of which is indicated by the genotype-dependent differences in tumour growth. The HIF-1 α -dependent growth arrest/apoptosis in response to hypoxia is reminiscent of the p53-mediated control of growth arrest/apoptosis in response to genotoxic or physical stresses^{11,12}. Furthermore, the HIF-1 α -dependent induction of p53, p21 and suppression of Bcl-2 during hypoxic growth arrest/apoptosis may reflect an interplay between these pathways. Indeed, hypoxia can stabilize p53 by binding HIF-1 α (ref. 24). The effects of losing HIF-1 α on the cell's behaviour confirm the importance of this molecular interaction. As HIF-1 α is activated in solid tumours, our results suggest a new general mechanism for tumour growth. □

Methods

HIF-1 α targeting, ES cells, embryoid bodies and tumour formation. The HIF-1 α targeting vector (pPNT.HIF-1 α ; Fig. 1a) contained a 5' flanking 5.0-kb *PstI*/*NheI* fragment upstream of exon 2, and a 3' flanking 4.8-kb *SphI* fragment downstream of exon 2. Culture and targeting of undifferentiated ES cells, including selection in high G418, northern blot and western blot analysis, and EMSA were done as described^{25,26}. For mRNA studies, ES cells were weaned off fibroblast feeders in the presence of leukaemia-inhibitory factor (LIF; ESGRO, GIBCO-BRL and Life Technologies). Cells at 75% confluency were cultured in normoxia (20% O₂)/normoglycaemia (25 mM glucose), in anoxia (<0.1% O₂; 6 h; BBL Gaspack 100 Anaerobic system, Beckton Dickinson), in hypoxia (0.5 or 2% O₂; 8–16 h) or in glucose-free medium (19 or 29 h) before mRNA analysis. Embryoid bodies (EBs) were generated by culturing ES cells (without feeders) in hanging drops (300 cells per 30 μ l) in medium without LIF for 3 d, and then on bacterial dishes for 2–8 weeks. EBs (3 weeks for *in situ* hybridization, 6 to 8 weeks for proliferation analysis) were stressed in hypoxia (1% O₂ for 24 h), or in glucose-free medium/anoxia (4 h; followed by 20 h recovery in normoxia/normoglycaemia). For ES-cell-derived tumour formation, 5×10^6 ES cells were subcutaneously injected into *Nu/Nu* mice.

Intravital microscopy, magnetic resonance imaging and flow. All methods for intravital microscopy of ES-cell-derived tumours in dorsal skin chambers, implanted in *Nu/Nu* mice including measurements of the velocity of red blood cells (V_{RBC}), blood-flow rates, vessel diameter and vessel density (defined as the total length of vessels per unit area; cm cm⁻²), and tissue pO_2 measurements were made at 15 to 17 days on tumours 8 to 10 mm in diameter (0.3 g) as described²³. Gradient echo MRI has been described²⁷. Vascular functionality ($b\Delta Y$) maps were derived from the average images, obtained by inhaling carbogen (95% O₂–5% CO₂) or air–CO₂ (95% air–5% CO₂); $b\Delta Y = (I_{\text{carbon}}/I_{\text{air-CO}_2})/TE \times C_{\text{MRI}}$, where ' b ' is the volume fraction of blood, ΔY is the change in fraction of oxyhaemoglobin, TE is time to echo (10 ms) and $C_{\text{MRI}} = 599 \text{ s}^{-1}$ at 4.7 T. Blood flow in tumours was measured using 15- μ m coloured microspheres as described²⁸.

Histology, immunostaining and *in situ* hybridization. All methods for histology, immunostaining, and *in situ* hybridization have been described^{26,29}. Vascular densities were quantified by counting the number of endothelial cords (consisting of two closely apposed endothelial cell layers without detectable lumen; <2 μ m wide; Fig. 2e), capillaries with a small lumen (diameter $7 \pm 5 \mu$ m; Fig. 2f), medium-sized vessels with a larger lumen ($30 \pm 3 \mu$ m versus $70 \pm 5 \mu$ m for medium and large diameter, respectively; Fig. 2g), or blood vessel lakes with the largest lumen ($90 \pm 14 \mu$ m, short axis versus $280 \pm 50 \mu$ m, long axis; Fig. 2h) per field (1.2 mm²) in 6 to 8 randomly chosen optical fields on 3 to 5 adjacent sections (320 μ m apart) per tumour. The fractional area of epithelioid cells (identified morphologically by their typical histological appearance and high cell density; Fig. S3n inset) was morphometrically quantified using the Quantimed Q600 imaging system.

Apoptosis and proliferation. For apoptosis studies, undifferentiated ES monolayers were cultured at 10,000 cells in 24-well dishes in normal ES cell medium containing 5% fetal calf serum for 24 h, and stressed for 24 h in the same medium without LIF under hypoxic or hypoglycaemic conditions. Apoptosis was measured using an ELISA (Boehringer Mannheim) for cytoplasmic histone-associated DNA fragments (mono- and oligonucleosomes), expressed as the number of oligonucleosomes per 10⁵ cells (undifferentiated ES cells) or per mg protein (tumours)³⁰. TUNEL-positive cells were visualized using an ApopTag kit (Oncor). Interleukin-1 β (3 ng ml⁻¹), interferon- γ (5 ng ml⁻¹), and tumour-necrosis factor- α (30 ng ml⁻¹) (all from R & D Systems) were added during normoxia/normoglycaemia for 24 h to induce hypoxia/hypoglycaemia-independent apoptosis. Oligonucleosomes (number per 10⁵ cells) and antigen levels of Bcl-2 (units per 10⁵ cells), p21 (units per 10⁵ cells) and p53 (pg per 10⁵ cells) were quantified by ELISA (Calbiochem). Apoptosis in hypoxic tumour zones was morphometrically quantified using the Quantimed Q600 imaging system by measuring the area of TUNEL-immunoreactive cells per hypoxic zone. Between 8 and 10 randomly chosen zones on 8 to 10 sections (320 μ m apart) were analysed and averaged per tumour. Western blotting of cell-cycle-control genes was done on whole-cell extracts using antibodies specific against p53, p21, Bcl-2 (all from Calbiochem), GADD153 (Santa Cruz Biotechnology) or β -actin (Sigma).

For analysis of proliferation of undifferentiated ES cell monolayers, ES cells (plated at 7,000 cells in 24-well dishes and serum-starved for 24 h) were stressed for 24 h in hypoxia (2% O₂; 24 h), incubated with 1 μ Ci ³H-thymidine for 2 h, and the incorporated ³H-thymidine counted³⁰. For analysis of proliferation in EBs, 6–8-week-old EBs were stressed under hypoxia (1% O₂, 24 h), or under hypoglycaemia/anoxia for 4 h, followed by a 20-h recovery in normoxia/normoglycaemia, in medium containing 5'-bromo-2',3'-deoxyuridine (BrdU; 30 μ M). *In situ* morphometric analysis of proliferation was done by measuring the area of all PCNA-immunoreactive cells on cross-section through each tumour (~150 mm²) and expressing it as a percentage of the tumour area analysed.

Received 2 April; accepted 12 June 1998.

- Bunn, H. F. & Poyton, R. O. Oxygen sensing and molecular adaptation to hypoxia. *Physiol. Rev.* **76**, 839–885 (1996).
- Wenger, R. H. & Gassmann, M. Oxygen(s) and the hypoxia-inducible factor-1. *Biol. Chem.* **378**, 609–616 (1997).
- Semenza, G. L. Transcriptional regulation by hypoxia-inducible factor-1. *Trends Cardiovasc. Med.* **6**, 151–157 (1996).
- Iyer, N. V. *et al.* Cellular and developmental control of O₂ homeostasis by hypoxia-inducible factor-1 α . *Genes Dev.* **12**, 149–162 (1998).
- Dor, Y. & Keshet, E. Ischemia-driven angiogenesis. *Trends Cardiovasc. Med.* **7**, 289–294 (1997).
- Gartel, A. L., Serfas, M. S. & Tyner, A. L. p21–negative regulator of the cell cycle. *Proc. Soc. Exp. Biol. Med.* **213**, 138–149 (1996).
- Toyoshima, H. & Hunter, T. p27, a novel inhibitor of G1 cyclin–Cdk protein kinase activity, is related to p21. *Cell* **78**, 67–74 (1994).
- Price, B. D. & Calderwood, S. K. Gadd45 and Gadd153 messenger RNA levels are increased during hypoxia and after exposure of cells to agents which elevate the levels of the glucose-regulated proteins. *Cancer Res.* **52**, 3814–3817 (1992).
- Little, E., Ramakrishnan, M., Roy, B., Gazit, G. & Lee, A. S. The glucose-regulated proteins (GRP78 and GRP94): functions, gene regulation, and applications. *Crit. Rev. Eukaryot. Gene Expr.* **4**, 1–18 (1994).
- Wood, S. M. *et al.* Selection and analysis of a mutant cell line defective in the hypoxia-inducible factor-1 α subunit. Characterization of HIF-1 α dependent and independent hypoxia-inducible gene expression. *J. Biol. Chem.* **273**, 8360–8367 (1998).
- Liebermann, D. A., Hoffman, B. & Steinman, R. A. Molecular controls of growth arrest and apoptosis: p53-dependent and independent pathways. *Oncogene* **11**, 199–210 (1995).
- Cox, L. S. Multiple pathways control cell growth and transformation: overlapping and independent activities of p53 and p21Cip1/WAF1/Sd1. *J. Pathol.* **183**, 134–140 (1997).
- Graeber, T. G. *et al.* Hypoxia induces accumulation of p53 protein, but activation of a G1-phase checkpoint by low-oxygen conditions is independent of p53 status. *Mol. Cell. Biol.* **14**, 6264–6277 (1994).
- Strasser, A., Huang, D. C. & Vaux, D. L. The role of the bcl-2/ced-9 gene family in cancer and general implications of defects in cell death control for tumorigenesis and resistance to chemotherapy. *Biochim. Biophys. Acta* **1333**, F151–F178 (1997).
- Maxwell, P. H. *et al.* Hypoxia-inducible factor-1 modulates gene expression in solid tumors and influences both angiogenesis and tumor growth. *Proc. Natl Acad. Sci. USA* **94**, 8104–8109 (1997).
- Maltepe, E., Schmidt, J. V., Baunoch, D., Bradfield, C. A. & Simon, C. M. Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT. *Nature* **386**, 403–407 (1997).
- Schmaltz, C., Harrigan Hardenbergh, P., Wells, A. & Fisher, D. E. Regulation of proliferation–survival decisions during tumor cell hypoxia. *Mol. Cell. Biol.* **18**, 2845–2854 (1998).
- Collins, M. K., Perkins, G. R., Rodriguez Tarduchy, G., Nieto, M. A. & Lopez Rivas, A. Growth factors as survival factors: regulation of apoptosis. *Bioessays* **16**, 133–138 (1994).
- Lin, Y. & Benchimol, S. Cytokines inhibit p53-mediated apoptosis but not p53-mediated G1 arrest. *Mol. Cell. Biol.* **15**, 6045–6054 (1995).
- Vaupel, P. The influence of tumor blood flow and microenvironmental factors on the efficacy of radiation, drugs and localized hyperthermia. *Klin. Padiatr.* **209**, 243–249 (1997).
- Yuan, F. *et al.* Time-dependent vascular regression and permeability changes in established human tumor xenografts induced by an anti-vascular endothelial growth factor/vascular permeability factor antibody. *Proc. Natl Acad. Sci. USA* **93**, 14765–14770 (1996).
- Graeber, T. G. *et al.* Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature* **379**, 88–91 (1996).
- Helmlinger, G., Yuan, F., Dellian, M. & Jain, R. K. Interstitial pH and pO_2 gradients in solid tumors *in vivo*: high-resolution measurements reveal a lack of correlation. *Nature Med.* **3**, 177–182 (1997).
- An, W. G. *et al.* Stabilization of wild-type p53 by hypoxia-inducible factor-1 α . *Nature* **392**, 405–408 (1998).
- Jiang, B. H., Rue, E., Wang, G. L., Roe, R. & Semenza, G. L. Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor-1. *J. Biol. Chem.* **271**, 17771–17778 (1996).
- Carmeliet, P. *et al.* Abnormal blood vessel development and lethality in embryos lacking a single vascular endothelial growth-factor allele. *Nature* **380**, 435–439 (1996).
- Abramovitch, R., Frenkel, D. & Neeman, M. Analysis of subcutaneous angiogenesis by gradient echo magnetic resonance imaging. *Magn. Reson. Med.* **39**, 813–824 (1998).
- Kobayashi, N., Kobayashi, K., Kouno, K., Horinaka, S. & Yagi, S. Effects of intra-arterial injection of colored microspheres on systemic hemodynamics and regional blood flow in rats. *Am. J. Physiol.* **266**, H1910–H1917 (1994).
- Evans, S. M. *et al.* Identification of hypoxia in cells and tissues of epigastric 9L rat glioma using EF5 (2-(2-nitro-1H-imidazol-1-yl)-N-(2,2,3,3,3-pentafluoropropyl) acetamide). *Br. J. Cancer* **72**, 875–882 (1995).
- Herbert, J. M. & Carmeliet, P. Involvement of u-PA in the anti-apoptotic activity of TGF β for vascular smooth muscle cells. *FEBS Lett.* **413**, 401–404 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank D. Livingston for monoclonal HIF-1 α antibodies; M. Lampugnani and E. Dejana for CD31 antibodies; E. M. Lord for ALK3.51 antibodies; G. Suske for Sp1 antibodies; S. Plaisance and G. Theilmeier for discussion and for their help; K. Bijnens, A. Bouché, I. Cornelissen, M. De Mol, E. Gils, B. Hermans, S. Jansen, L. Kieckens, A. Manderveld, T. Vancotsem, A. Vandenhoek, A. Van den Boomen, P. Van Wesemael, S. Wyns (Leuven), A. Itin, I. Lamarche, P. Rogon, Y. Chen, J. Kahn and T. Gohongi for technical assistance; and M. Deprez for artwork. This work was supported in part by a NCI-01G grant to R.J. and D.E.

Correspondence and requests for materials should be addressed to P.C. (e-mail: Peter.Carmeliet@med.kuleuven.ac.be).

Role of citron kinase as a target of the small GTPase Rho in cytokinesis

Pascal Madaule, Masatoshi Eda, Naoki Watanabe, Kazuko Fujisawa, Toshiyuki Matsuoka, Haruhiko Bito, Toshimasa Ishizaki & Shuh Narumiya

Department of Pharmacology, Kyoto University Faculty of Medicine, Sakyo-ku, Kyoto 606-8315, Japan

During mitosis, a ring containing actin and myosin appears beneath the equatorial surface of animal cells. This ring then contracts, forms a cleavage furrow and divides the cell¹⁻³, a step known as cytokinesis. The two daughter cells often remain connected by an intercellular bridge which contains a refringent structure known as the midbody^{4,5}. How the appearance of this ring is regulated is unclear, although the small GTPase Rho, which controls the formation of actin structures^{6,7}, is known to be essential⁸⁻¹⁰. Protein kinases are also thought to participate in cytokinesis^{1-3,11,12}. We now show that a splice variant of a Rho target protein, named citron¹³, contains a protein kinase domain that is related to the Rho-associated kinases ROCK¹⁴ and ROK¹⁵⁻¹⁷, which regulate myosin-based contractility¹⁸⁻²¹. Citron kinase localizes to the cleavage furrow and midbody of HeLa cells; Rho is also localized in the midbody. We find that overexpression of *citron* mutants results in the production of multinucleate cells and that a kinase-active mutant causes abnormal contraction during cytokinesis. We propose that citron kinase regulates cytokinesis at a step after Rho in the contractile process.

We previously identified a partner of the active form of Rho. This 183K protein, termed citron¹³, resembles the ROCK family of kinases¹⁴⁻¹⁷ in its overall domain structure but lacks a kinase domain. Using a cloning approach based on the polymerase chain reaction (PCR), we have now identified a splice variant of citron with an alternative amino terminus. This N-terminal extension contains a protein kinase domain that has ~50% sequence identity to the sequences of ROCK, ROK, myotonic dystrophy protein kinase (MDPK) (Fig. 1a, b) and the Cdc42 effector known as MRCK or GEK^{22,23}. Three carboxy-terminal truncation mutants of citron, $\Delta 1$, $\Delta 2$ and $\Delta 3$, were then produced in HeLa cells, and subjected to an immuno-kinase assay as described¹⁴. The mutants all phosphorylated histones to a similar extent (Fig. 1c). We then investigated the tissue distribution of the citron isoforms with and without the kinase domain. Northern blot analysis indicated that messenger RNA encoding the non-kinase form is abundant only in brain, whereas the kinase form is widely expressed (data not shown). We refer to the kinase form as citron-K and the non-kinase form as citron-N (for neuron or non-kinase). Antibodies against the Rho-binding region of citron were generated in three rabbits and used for immunoblot analysis (Fig. 1d). An immunoreactive protein with a mobility identical to that of citron-N was detected only in the brain, whereas a band corresponding to citron-K was detected only in the testis and HeLa cells. The antisera did not crossreact with ROCK or ROK (data not shown). Thus, the predominant isoform of citron in testis and HeLa cells is citron-K.

The localization of citron-K was then determined by immunofluorescence analysis with untransfected HeLa cells. The three citron-specific antisera yielded identical results. In dividing cells, citron-K appeared to be associated with the cleavage furrow (Fig. 2a-c). In addition, a bright doughnut-shaped dot was often observed between two interphase cells (Fig. 2c): double immunostaining revealed that this dot lies between two thick bundles of

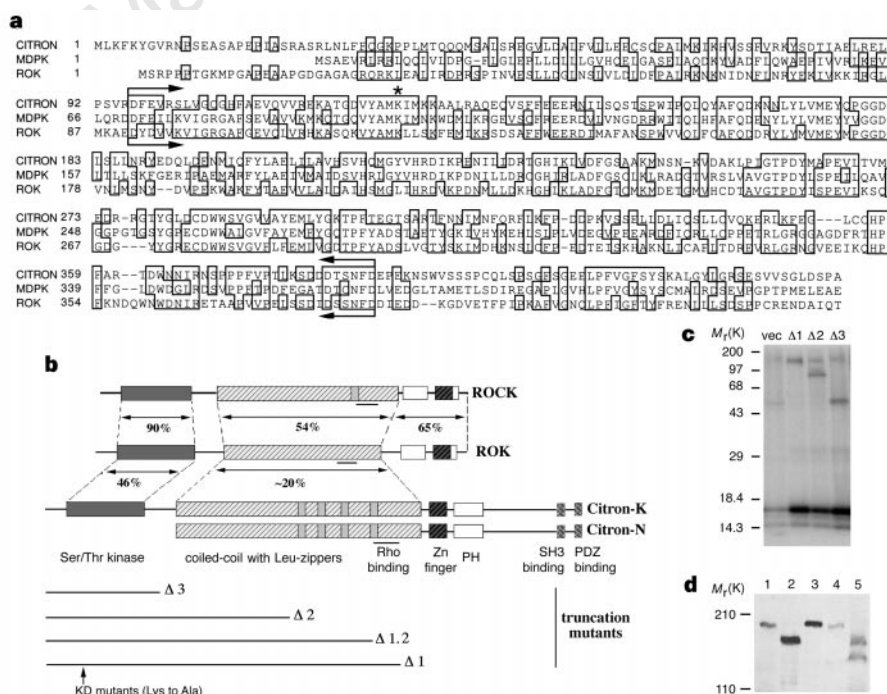


Figure 1 Structure and expression of citron kinase. **a**, The predicted amino-acid sequence of the kinase domain of citron is compared with the N-terminal regions of ROK and of MDPK. The kinase domains are delineated by arrows. The star indicates the lysine that is mutated to alanine in the citron-K kinase-defective (KD) mutants. **b**, Domain structure of the ROCK family of kinases. Percentage identities to ROK are indicated. Truncation mutants of citron-K and the kinase-defective mutation are shown. **c**, Kinase assay of recombinant citron-K. HeLa cells were

transfected with the indicated constructs. Vec: vector alone. Expressed proteins were assayed for kinase activity with histones ($M_r \sim 15K$) and $[\gamma\text{-}^{32}P]\text{ATP}$ as substrates. **d**, Immunoblot of citron-K and citron-N. Cell lysates (15 to 30 μg) (lanes 1, 4, 5) or 2 μg (lanes 2, 3) of proteins were analysed with antiserum specific for citron. Lanes: 1, untransfected HeLa cells; 2, HeLa cells transfected with citron-N; 3, HeLa cells transfected with citron-K; 4, testis; 5, brain. The lower band in the brain sample was not detected with other antisera against citron.

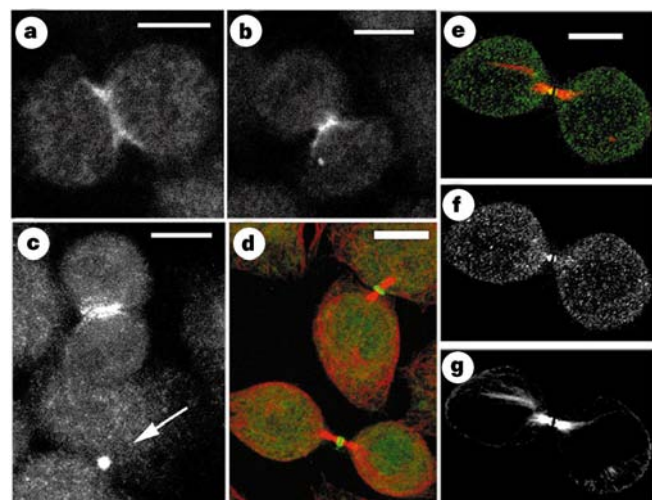


Figure 2 Localization by immunofluorescence of endogenous citron-K and RhoA in untransfected HeLa cells. **a–c**, Staining with a rabbit antiserum specific for citron. Arrow indicates a midbody in interphase cells. **d**, Double staining for tubulin (red) and citron (green). **e**, Double staining for tubulin (red) and RhoA (green). **f**, Same as **e**, but RhoA stain only. **g**, Same as **e**, but tubulin stain only. Scale bars, 10 μ m.

tubulin (Fig. 2d), in a pattern typical of the intercellular bridge. Citron is located in the middle of the bridge, a region known as the midbody.

To confirm the localization of citron-K, we transfected HeLa cells with plasmids encoding Myc-tagged citron-K, either as the full-length protein or as C-terminal truncation mutants. Citron was visualized with antibodies against Myc at 24 h after transfection. Myc-tagged citron-K (full-length) was detected in the cleavage furrow of dividing cells as well as in the midbody (Fig. 3a, b). The $\Delta 1$ protein also localized to the cleavage furrow, from the early stages until almost the completion of cytokinesis (Fig. 3c–e). Colocalization of endogenous Rho with overexpressed citron was apparent in the cleavage furrow and the midbody (Fig. 3f–k), indicating that citron binds to Rho *in vivo*. In untransfected cells, Rho was detected in the midbody at the tip of the microtubules (Fig. 2e–g), but not in the cleavage furrow under our experimental conditions.

Most $\Delta 1$ and full-length citron-K appeared as large dots, especially during interphase (Figs 3, 4a). These structures could be aggregates or they could represent an enlarged membrane compartment where citron accumulates; the $\Delta 1$ protein was easily solubilized with 0.05% Triton-X100. On the other hand, $\Delta 2$ and $\Delta 3$ proteins (Fig. 4b) showed a diffuse cytosolic staining in both mitotic and interphase cells, although a small amount of $\Delta 2$ was associated with the midbody. These observations suggest that the C-terminal portion of the coiled-coil domain present in $\Delta 1$ but not in $\Delta 2$ is important for localization. Because yeast two-hybrid analysis indicated that $\Delta 1$ binds Rho-GTP, we deleted the whole Rho-binding region from $\Delta 1$ (Fig. 1b). The resulting protein, $\Delta 1.2$, showed an essentially cytosolic distribution (data not shown), indicating that the Rho-binding region of citron is required for localization to the cleavage furrow and the midbody.

To verify that citron acts in cytokinesis, we transfected HeLa cells with complementary DNA encoding the various citron mutants and examined the cells 3 days later, a period sufficient for two or three cell divisions. Nearly all of the cells that strongly expressed $\Delta 1$ had become multinucleated (Fig. 4a). Thus, overexpression of the truncated citron-K mutant specifically blocked cytokinesis without markedly affecting cell viability, cell-cycle progression or nuclear division. $\Delta 2$ also produced the multinucleate phenotype in ~50%

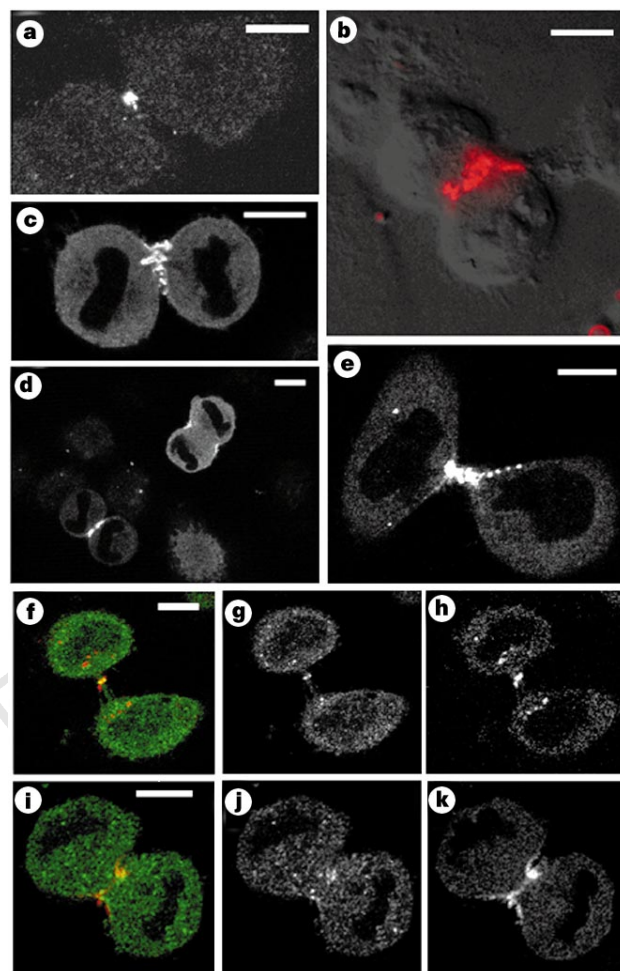


Figure 3 Localization of recombinant citron-K. HeLa cells were transfected with cDNAs encoding full-length citron-K (**a**, **b**), $\Delta 1$ (**c**–**e**) or $\Delta 1$ KD (**f**–**k**). After 24 h, cells were processed for immunofluorescence analysis. Recombinant proteins were detected with antibody against Myc. In **b**, cells were examined with a combination of transmitted light (grey) and fluorescence (red). **f**, **i**, Double staining for recombinant Myc-citron (red) and endogenous RhoA (green). **g**, **j**, Same as **f** and **i**, but RhoA stain only. **h**, **k**, Same as **f** and **i**, by Myc stain only. Scale bars, 10 μ m.

of the strongly expressing cells (Fig. 4b). Expression of full-length citron-K or $\Delta 3$ resulted in the formation of only a few multinucleate cells.

We next expressed citron devoid of kinase activity. We prepared kinase-defective mutants (denoted as KD) by replacing the conserved lysine at position 126, in the ATP-binding site, with alanine. The cells were analysed 2 days after transfection. For citron-N, $\Delta 1$ KD, $\Delta 1.2$ KD, and $\Delta 2$ KD, the proportion of highly expressing cells that became multinucleate was 41, 11, 8 and 9%, respectively, compared with 1.8% in control cells. Furthermore, flow-cytometry analysis (Fig. 4c) showed that the proportion of cells with a DNA content of 4N among control cells and those expressing $\Delta 1$ (kinase active) or $\Delta 1.2$ KD was 23.5 ± 0.6 , 45.0 ± 0.8 and $34.8 \pm 1.3\%$, respectively (means $\pm$ s.e.m. from 3 independent experiments; $P < 0.001$ for $\Delta 1.2$ KD versus control and $P < 0.0001$ for $\Delta 1$ versus control according to the *t*-test). Thus, kinase activity is important in the phenotype of $\Delta 1$ and $\Delta 2$. Mutants lacking kinase activity, however, retain a substantial effect on cytokinesis.

The $\Delta 1$ phenotype was analysed by time-lapse videomicroscopy. Cytokinesis was initiated normally at the end of telophase, but the transfected cells consistently failed to establish the intercellular bridge and reversed their cytokinesis within 1 h. In ~20% of the

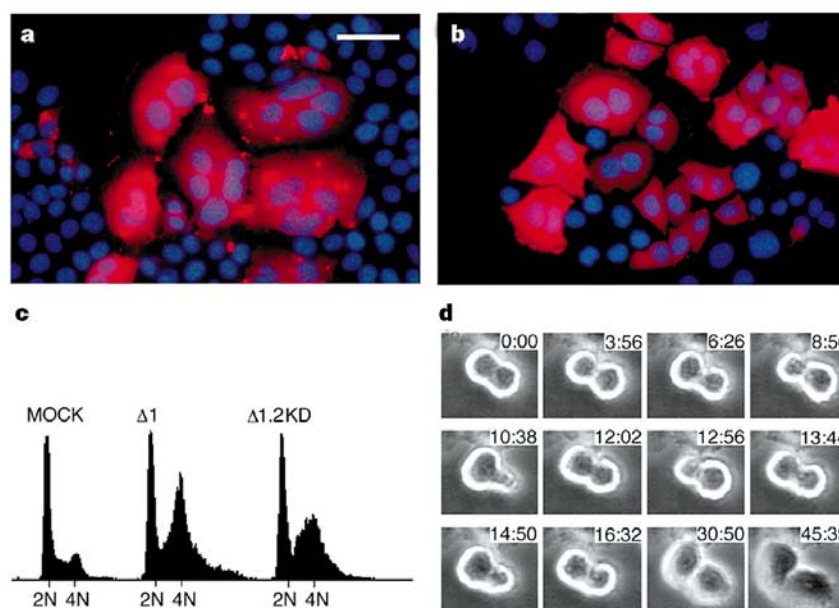


Figure 4 Effects of citron-K mutants on cytokinesis. HeLa cells were transfected with cDNAs encoding $\Delta 1$ (**a**) or $\Delta 2$ (**b**) and processed for immunofluorescence after 3 days. Nuclei were stained with DAPI (blue), and recombinant proteins with antibody against Myc (red). Scale bar, 25 μm . **c**, Flow cytometry of the DNA contents of cells expressing $\Delta 1$ or $\Delta 1.2\text{KD}$, or mock-transfectants. **d**, Videomicroscopy of cells expressing $\Delta 1$. HeLa cells were co-transfected with $\Delta 1$ cDNA (1 μg)

and pEGFP-C1 (0.1 μg), a plasmid encoding green fluorescent protein (GFP). One day later, transfected cells identified by GFP were monitored by time-lapse phase-contrast videomicroscopy. Images are shown at selected time points (min:sec). Cytokinesis failed at 45 min. The full movie is available as Supplementary Information.

cells, the cleavage furrow prematurely ceased contracting and cytokinesis reversed immediately. The most marked feature of this phenotype was its abnormal contractile movement which was apparent in $\sim 50\%$ of cells recorded (Fig. 4d). In these cases, one of the prospective daughter cells contracted unequally during cytokinesis, transferring most of its contents to its partner, which then itself contracted and pushed back the transferred material. These push-pull movements were rapid, each cycle requiring ~ 1 min, and occurred repeatedly. Such abnormal contractile movements were not observed in cells expressing $\Delta 1\text{KD}$. Thus, the abnormal contractions require the kinase activity of $\Delta 1$. We also stained cells transfected with various citron-K mutants for F-actin and saw a normal actin pattern during mitosis. Rho, however, is required for actin accumulation at the cleavage furrow^{8–10}. Citron-K may work in the contractile process while another Rho target regulates actin accumulation: p140mDia could be a candidate for the latter as this protein induces actin polymerization²⁴ and its *Drosophila* homologue is required for cytokinesis²⁵.

The structural similarity of citron to ROCK suggests a similarity in function. ROCK stimulates myosin-based contractility²¹ and can induce a massive formation of stress fibres^{18,19}. In contrast, overexpression of citron-K had, at most, a marginal effect on stress-fibre formation. Conversely, overexpression of ROCK mutants did not result in the formation of a significant number of multinucleate cells (data not shown). Furthermore, Y-27632, a specific ROCK inhibitor²⁶ which blocks stress-fibre formation at 10 μM , did not affect cytokinesis in HeLa cells at concentrations lower than 100 μM . In the kinase assay, citron-K was 30 times less sensitive to Y-27632 than ROCK (data not shown). Thus, ROCK does not appear to be essential for cytokinesis. On the other hand, the localization of citron to the cleavage furrow and the abnormal movements induced by $\Delta 1$ during cell division are indicative of a role for this Rho effector in the contractile process of cytokinesis. □

Methods

Cloning of the kinase domain of citron. We previously isolated partial cDNAs corresponding to transcripts generated by alternative splicing at codon 8 of citron-N and encoding an N-terminal extension¹³. The longest clone extends 283 bp upstream of this splice point and displayed homology to the kinase domain of ROCK. Preliminary RT-PCR reactions allowed determination of the sequence for the kinase domain of citron, which was complete, as confirmed by 5'-RACE. Additional RT-PCR confirmed that this sequence entirely derives from citron. A final RT-PCR was done using Pfu polymerase (Stratagene), poly(A)⁺ RNA from PC12 cells, and the primers AGTCGGTAGCGGAGAGA and TCTTTTCCATGGAGCTAACC. After 25 cycles, a single 1.4-kb band was observed. The products of two independent reactions were subcloned into pBluescript (Stratagene) (yielding of pKIN) and each was entirely sequenced on both strands. These products were identical to each other.

Expression constructs. A Myc epitope tag, a *KpnI* cloning site, and stop codons were added to the vector pCAG-GS in which expression is controlled by the β -actin promoter²⁷, yielding pCAG-MycSTOP. A *KpnI* site was created just upstream of the putative initiation codon of the pKIN insert, generating the $\Delta 3$ (448 codons) truncation mutant. The full-length citron-K (2,013 codons) was reconstructed with the use of a unique *BglII* site present in the 230-bp overlap between the pKIN insert and our previous long citron cDNA¹³. Digestion of the full-length citron-K cDNA with *NheI* or *HindIII* generated $\Delta 1$ (1,247 codons) and $\Delta 2$ (788 codons) truncation mutants, respectively. A stop codon, as well as the mutation at codon 126 (Lys to Ala), were generated by PCR with mutagenic primers, allowing construction of $\Delta 1.2$ (1,123 codons) and the kinase-defective alleles. Myc-citron-N cDNA has been described¹³. These constructs were introduced into pCAG-MycSTOP. Cells were transfected by the use of Lipofectamine (GIBCO-BRL). Constructs for expression of ROCK mutants have been described¹⁸.

Antibodies, immunoblot and immunofluorescence. The His-tagged Rho-binding region of citron was prepared from bacteria as described¹³ and injected into rabbits. Immunoblots were incubated with these rabbit antisera at a dilution of 1/1,000, and processed with the ECL system (Amersham). For immunofluorescence, antisera against citron were diluted 1/300 and fluores-

cent secondary antibodies were diluted 1/200, with the exception of that shown in Fig. 3f–k, where Texas-red-labelled antibodies were diluted 1/700. Figures 2e–g and 3a, c–k represent individual confocal sections. The mouse monoclonal antibodies to tubulin and Myc were TUB2.1 (Sigma) and 9E10, respectively. Rabbit antibodies to RhoA (Santa Cruz) were diluted 1/100 times. Fluorescence intensity in individual cells was measured by a camera attached to the fluorescence microscope (Axiophot, Zeiss). A transfected cell was considered as highly expressing a protein if it required an exposure time of less than 10 s. This value was observed in ~20% of the transfected cells. The rate of spontaneous cytokinesis failure, 1.8%, was estimated with control cells co-transfected with 1 µg pCAG-MycSTOP and 100 ng pEGFP-C1 (Clontech).

Flow cytometry. Two days after transfection, cells were trypsinized, fixed with 0.5% formaldehyde and permeabilized with 0.1% Triton X-100. They were stained for recombinant citron with antibodies to Myc and FITC-conjugated secondary antibodies, and for DNA with propidium iodide. The stained cells were subjected to flow cytometry on a FACScalibur (Becton Dickinson). Citron-transfected cells positive for FITC and mock-transfected cells were analysed for DNA content.

Effect of Y-27632 on cytokinesis. HeLa cells were synchronized with thymidine and nocodazole as described²⁸. After this procedure, ~60% of the cells appeared to have arrested in the M phase of the cell cycle. Y-27632 was added to the culture medium at various concentrations for the last 30 min of the nocodazole incubation. Cells were then washed three times and incubated with medium containing Y-27632 for 3 h. The percentage of mitotic cells that became binucleate was determined visually.

Received 5 May; accepted 17 June 1998.

- Mabuchi, I. Biochemical aspects of cytokinesis. *Int. Rev. Cytol.* **101**, 175–213 (1986).
- Satterwhite, L. L. & Pollard, T. D. Cytokinesis. *Curr. Opin. Cell Biol.* **4**, 43–52 (1992).
- Fishkind, D. J. & Wang, Y.-L. New horizons for cytokinesis. *Curr. Opin. Cell Biol.* **7**, 23–31 (1995).
- Byer, B. & Abramson, D. H. Cytokinesis in HeLa: post-telophase delay and microtubule-associated motility. *Protoplasma* **66**, 413–435 (1968).
- Mullins, J. M. & Biesse, J. J. Terminal phase of cytokinesis in D-98S cells. *J. Cell Biol.* **73**, 672–684 (1977).
- Machesky, L. M. & Hall, A. Rho: a connection between membrane receptor signalling and the cytoskeleton. *Trends Cell Biol.* **6**, 304–310 (1996).
- Narumiya, S. The small GTPase Rho: cellular functions and signal transduction. *J. Biochem. (Tokyo)* **120**, 215–228 (1996).
- Mabuchi, I. et al. A rho-like protein is involved in the organisation of the contractile ring in dividing sand dollar eggs. *Zygote* **1**, 325–331 (1993).
- Kishi, K., Sasaki, T., Kuroda, S., Itoh, T. & Takai, Y. Regulation of cytoplasmic division of *Xenopus* embryo by rho p21 and its inhibitory GDP/GTP exchange protein (rho GDI). *J. Cell Biol.* **120**, 1187–1195 (1993).
- Drechsel, D. N., Hyman, A. A., Hall, A. & Glotzer, M. A requirement for Rho and Cdc42 during cytokinesis in *Xenopus* embryos. *Curr. Biol.* **7**, 12–23 (1996).
- Terada, Y. et al. AIM-1: a mammalian midbody associated protein required for cytokinesis. *EMBO J.* **17**, 667–676 (1998).
- Sekimata, M. et al. Detection of protein kinase activity specifically activated at metaphase–anaphase transition. *J. Cell Biol.* **132**, 635–641 (1996).
- Madaule, P. et al. A novel partner for the GTP-bound forms of rho and rac. *FEBS Lett.* **377**, 243–248 (1995).
- Ishizaki, T. et al. The small GTP-binding protein Rho binds to and activates a 160 kDa Ser/Thr protein kinase homologous to myotonic dystrophy kinase. *EMBO J.* **15**, 1885–1893 (1996).
- Leung, T., Manser, E., Tan, L. & Lim, L. A novel serine/threonine kinase binding the Ras-related RhoA GTPase which translocates the kinase to peripheral membranes. *J. Biol. Chem.* **270**, 29051–29054 (1995).
- Matsui, T. et al. Rho-associated kinase, a novel serine/threonine kinase, as a putative target for the small GTP binding protein Rho. *EMBO J.* **15**, 2208–2216 (1996).
- Nakagawa, O. et al. ROCK-1 and ROCK-2: two isoforms of Rho-associated coiled-coil forming protein serine/threonine kinase in mice. *FEBS Lett.* **392**, 189–193 (1996).
- Ishizaki, T. et al. p160ROCK, a Rho-associated coiled-coil forming protein kinase, works downstream of Rho and induces focal adhesions. *FEBS Lett.* **404**, 118–124 (1997).
- Leung, T., Chen, X.-Q., Manser, E. & Lim, L. The p160 RhoA-binding kinase ROKα is a member of a kinase family and is involved in the reorganization of the cytoskeleton. *Mol. Cell. Biol.* **16**, 5313–5327 (1996).
- Kimura, K. et al. Regulation of myosin phosphatase by Rho and Rho-associated kinase (Rho-kinase). *Science* **273**, 245–248 (1996).
- Kureishi, Y. et al. Rho-associated kinase directly induces smooth muscle contraction through myosin light chain phosphorylation. *J. Biol. Chem.* **272**, 12257–12260 (1997).
- Luo, L. et al. Genghis Khan (Gek) as a putative effector for *Drosophila* Cdc42 and regulator of actin polymerization. *Proc. Natl Acad. Sci. USA* **94**, 12963–12968 (1997).
- Leung, T., Chen, X.-Q., Tan, L., Manser, E. & Lim, L. Myotonic dystrophy kinase-related Cdc42-binding kinase acts as a Cdc42 effector in promoting cytoskeletal reorganization. *Mol. Cell. Biol.* **18**, 130–140 (1998).
- Watanabe, N. et al. p140mDia, a mammalian homolog of *Drosophila* diaphanous, is a target protein for Rho small GTPase and is a ligand for profilin. *EMBO J.* **16**, 3044–3056 (1997).
- Castrillon, D. & Wasserman, S. diaphanous is required for cytokinesis in *Drosophila* and shares domains of similarity with the products of the limb deformity gene. *Development* **120**, 3367–3377 (1994).
- Uehata, M. et al. Calcium sensitization of smooth muscle mediated by a Rho-associated protein kinase in hypertension. *Nature* **389**, 990–994 (1997).
- Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–200 (1991).
- Andreassen, P., Palmer, D., Wener, M. & Margolis, R. Telophase disc: a new mammalian mitotic organelle that bisects telophase cells with a possible function in cytokinesis. *J. Cell Sci.* **99**, 523–534 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>).

Acknowledgements. We thank S. Yonemura and I. Mabuchi for critically reading the manuscript and H. A. Popiel for editing it; H. Fuyuhiko and K. Nonomura for technical assistance; and T. Arai, H. Nose and K. Okuyama for secretarial assistance. This work was supported in part by a grant from the Ministry of Education, Sciences, Sports and Culture of Japan and by a grant for the Human Frontier Science Program. P.M. was supported in part by the Ciba–Geigy Foundation.

Correspondence and requests for materials should be addressed to S.N. (e-mail: snaru@mfour.med.kyoto-u.ac.jp). The nucleotide sequences encoding citron kinase have been deposited in GenBank under accession numbers AF070065 and AF070066.

EEA1 links PI(3)K function to Rab5 regulation of endosome fusion

Anne Simonsen^{*†}, Roger Lippé^{†‡#}, Savvas Christoforidis^{‡#}, Jean-Michel Gaullier^{*}, Andreas Brech[§], Judy Callaghan^{||}, Ban-Hock Toh^{||}, Carol Murphy^{‡¶#}, Marino Zerial^{‡#} & Harald Stenmark^{*}

^{*} Department of Biochemistry, The Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway

[‡] EMBL, Meyerhofstrasse 1, D-69012 Heidelberg, Germany

[§] EM-unit, Institute of Biology, PO Box 1050, Blindern, N-0316 Oslo, Norway

^{||} Department of Pathology and Immunology, Monash Medical School, Prahran, Victoria 3181, Australia

[#] Max Planck Institute for Molecular Cell Biology and Genetics, Dresden, 01307, Germany

[†] These authors contributed equally to this work

GTPases and lipid kinases regulate membrane traffic along the endocytic pathway by mechanisms that are not completely understood^{1–4}. Fusion between early endosomes requires phosphatidylinositol-3-OH kinase (PI(3)K) activity^{5–7} as well as the small GTPase Rab5 (ref. 8). Excess Rab5–GTP complex restores endosome fusion when PI(3)K is inhibited^{5,9}. Here we identify the early-endosomal autoantigen EEA1 (refs 10–12) which binds the PI(3)K product phosphatidylinositol-3-phosphate, as a new Rab5 effector that is required for endosome fusion. The association of EEA1 with the endosomal membrane requires Rab5–GTP and PI(3)K activity, and excess Rab5–GTP stabilizes the membrane association of EEA1 even when PI(3)K is inhibited. The identification of EEA1 as a direct Rab5 effector provides a molecular link between PI(3)K and Rab5, and its restricted distribution to early endosomes¹⁰ indicates that EEA1 may confer directionality to Rab5-dependent endocytic transport.

EEA1 binds to phosphatidylinositol-3-phosphate (PtdIns(3)-P)^{11,12} and localizes together with Rab5 on endosomes^{10,12,13}. Because a GTP-hydrolysis-deficient mutant of Rab5 can counteract the effect of the PI(3)K inhibitor wortmannin on endosome fusion^{5,9}, we determined whether EEA1 might physically interact with Rab5. Using the yeast two-hybrid system (Table 1), we first found that wild-type Rab5 and a mutant with preferential affinity for GDP, Rab5^{S34N}, in which serine 34 is converted to asparagine, showed minimal interaction with EEA1. Remarkably, the GTPase-deficient mutant Rab5^{Q79L} interacted strongly with EEA1, indicating that the GTP-bound form of Rab5 may recognize EEA1. We detected no interaction between EEA1 and the corresponding GTPase-deficient mutants of other Rab proteins (Rab3A^{Q81L}, Rab3C^{Q81L}, Rab4A^{Q67L}, Rab4B^{Q67L}, Rab6^{Q72L}, Rab7^{Q67L} and Rab11^{Q70L}) (Table 1), indicating that EEA1 interacts specifically with Rab5.

A deletion analysis showed that both amino- and carboxy-terminal EEA1 fragments (EEA1-NT and EEA1-CT, respectively; Fig. 1a) can interact with Rab5^{Q79L} in the two-hybrid system (Table 1). We first found that EEA1-NT (Fig. 1a) interacts with

[¶] Present address: University of Ioannina Medical School, 45110 Ioannina, Greece.

Table 1 Interactions between EEA1 and Rab proteins in the two-hybrid system

	EEA1	
Rab5	0.53 ± 0.21	
Rab5 ^{Q79L}	101 ± 33.3	
Rab5 ^{S34N}	0.15 ± 0.05	
Rab3A ^{Q81L}	0.07 ± 0.01	
Rab3C ^{Q81L}	0.26 ± 0.13	
Rab4A ^{Q67L}	0.12 ± 0.01	
Rab4B ^{Q67L}	0.05 ± 0.04	
Rab6 ^{Q72L}	0.04 ± 0.01	
Rab7 ^{Q67L}	0.06 ± 0.02	
Rab11 ^{Q70L}	0.07 ± 0.01	
	Rab5 ^{Q79L}	Rab5 ^{S34N}
EEA1 ^{1,257-1,411} *	25.6 ± 0.9	0.11 ± 0.07
EEA1 ^{1,277-1,411} *	32.0 ± 11.2	0.09 ± 0.04
EEA1 ^{1,325-1,411} *	0.77 ± 0.02	0.11 ± 0.06
EEA1 ^{1,277-1,357} *	15.4 ± 1.5	0.21 ± 0.15
EEA1 ^{1,277-1,348} *	4.16 ± 0.51	0.21 ± 0.06
EEA1 ^{1,307-1,357} *	0.13 ± 0.02	0.14 ± 0.03
EEA1 ¹⁻²⁰⁹	512 ± 43.1	0.17 ± 0.02
EEA1 ¹⁻²⁰⁹ ^{C46S}	0.08 ± 0.01	-
EEA1 ¹⁻¹⁵³	95.2 ± 13.6	0.08 ± 0.01
EEA1 ¹⁻⁹¹	17.4 ± 0.97	0.07 ± 0.01

Results were obtained from yeast two-hybrid assays by measurement of *lacZ* reporter activity. L40 reporter yeast cells were transformed with Rab constructs in plasmid pLexA, and EEA1 constructs in plasmid pGAD, or vice versa where indicated with an asterisk. The cells were grown to A₆₀₀ values of roughly 1.0 in synthetic medium lacking tryptophan and leucine. β-Galactosidase activity was then measured using *o*-nitrophenyl-β-D-galactopyranoside as a substrate. β-Galactosidase activities (in relative units) are presented as mean values ± deviations between duplicate transformants.

Rab5^{Q79L} and not with Rab5^{S34N}. This interaction presumably involves the C₂H₂ zinc-finger motif of EEA1, as a point mutation in this motif (producing EEA1-NT^{C46S}) abolished the interaction with Rab5^{Q79L}, and a small fragment containing the zinc finger (EEA1₁₋₉₁) retained the ability to interact. We further found that, like EEA1-NT, EEA1-CT, which contains a PtdIns(3)P-binding FYVE finger¹², interacts with Rab5^{Q79L} but not with Rab5^{S34N}. The entire FYVE finger could be removed with only a weak reduction in Rab5^{Q79L} binding (Table 1). The minimal Rab5-binding region of EEA1-CT resides within residues 1,277–1,348, which constitute the domain immediately N-terminal of the FYVE finger. The region encompassing amino-acid residues 1,306–1,340 of EEA1 is 29% identical and 43% similar to the Rab5-binding region¹⁴ of the Rab5 effector Rabaptin-5 (results not shown), indicating that the Rab5-binding regions of EEA1-CT and Rabaptin-5 may be structurally related.

We next confirmed the Rab5–EEA1 interactions biochemically. EEA1 from bovine brain cytosol bound specifically to immobilized Rab5 fused to maltose-binding protein (MBP–Rab5). We detected binding only in the presence of the non-hydrolysable GTP analogue GTP-γS, and not in the presence of GDP or the absence of nucleotide (Fig. 1b). In agreement with the two-hybrid data, we detected only background-level interaction of EEA1 with MBP–Rab4 and with MBP–Rab6, even in the presence of GTP-γS. Moreover, recombinant His₆–EEA1 protein bound to an immobilized glutathione *S*-transferase (GST)–Rab5 fusion protein in the GTPγS-bound form, but not in the GDP-bound form (Fig. 1c), showing that the interaction is direct and independent of other cytosolic factors. Finally, we also demonstrated GTPγS-specific interactions between GST–Rab5 and MBP–EEA1-NT (Fig. 1d) or MBP–EEA1-CT (Fig. 1e), confirming the results from the two-hybrid assays.

Next we determined whether EEA1 functions in early-endosome fusion. Expression of Rab5^{Q79L} leads to the expansion of early endosomes in BHK cells because of increased fusion activity (Fig. 2a)¹⁵. Although the expression of Myc-epitope-tagged EEA1-NT with Rab5^{Q79L} had no effect (results not shown), the coexpression of Myc-tagged EEA1-CT (Fig. 2b) counteracted the activity of the Rab5 mutant, resulting in the clustering of small endocytic structures (Fig. 2b, inset) often located in the periphery of the cell. In contrast, the FYVE-finger mutant EEA1-CT^{C1358S}, which does not

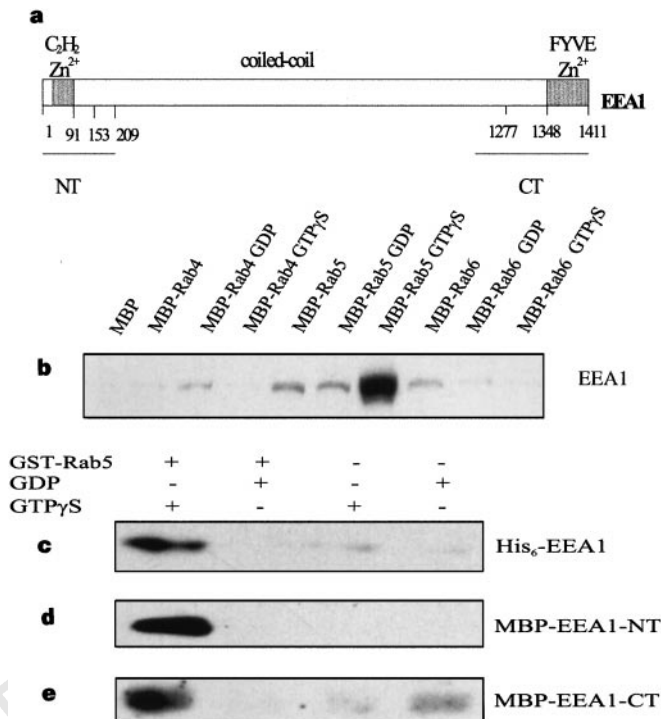


Figure 1 Biochemical interaction between EEA1 and Rab5. **a**, Representation of wild-type EEA1, with the C₂H₂ and FYVE zinc fingers indicated. The EEA1 constructs are indicated by their amino-acid numbers. The point mutants used were C46S in the C₂H₂ finger and C1358S in the FYVE finger. We use the terms EEA1-NT for EEA1₁₋₂₀₉ and EEA1-CT for EEA1_{1,257-1,411}. **b**, MBP–Rab4, MBP–Rab5 or MBP–Rab6, immobilized on Affi-Gel, were either nucleotide-free or loaded with GDP or GTP-γS and then incubated with bovine brain cytosol to study their binding to EEA1. **c–e**, Glutathione–agarose beads, with or without bound GST–Rab5, were incubated with GDP or GTP-γS and then mixed with recombinant His₆–EEA1 (**c**), MBP–EEA1-NT (**d**), or MBP–EEA1-CT (**e**). Bound proteins were eluted with reduced glutathione. Bound proteins in **b** and eluted fractions in **c–e** were analysed by SDS–PAGE and immunoblotting with anti-EEA1 antiserum (**b–d**) or anti-MBP antibodies (**e**).

bind to endosomes¹³, did not affect the expansion of early endosomes (Fig. 2c), indicating that the effect of EEA1-CT requires its binding to endosomes. We studied the effect of EEA1-CT on early endosomes under normal conditions (that is, in the absence of the stimulatory Rab5 mutant) by electron microscopy. Cells expressing EEA1-CT contained many small, early-endocytic profiles (Fig. 2d), which were significantly smaller than the early endosomes of untransfected cells (Fig. 2e). These results indicate that the expression of EEA1-CT affects the size of normal as well as Rab5^{Q79L}-containing early endosomes.

To determine whether EEA1-CT inhibits early-endosome fusion directly, we used an *in vitro* endosome fusion assay based on the detection of complexes formed between ligands internalized into two separate early endosome populations^{8,16}. We prepared cytosol from baby hamster kidney (BHK) cells overexpressing EEA1-CT or EEA1-CT^{C1358S} and tested this cytosol in the fusion assay in the presence of cytosol from HeLa cells. The fusion activity was cytosol-dependent (Fig. 3a). We detected 60% inhibition of fusion when early endosomes were incubated in the presence of cytosol containing EEA1-CT, but not in the presence of cytosol containing EEA1-CT^{C1358S}. An anti-EEA1 serum, but not its corresponding preimmune serum, inhibited early-endosome fusion by ~80% when added directly to the fusion reaction (Fig. 3b). Addition of the peptide against which the antiserum had been raised completely rescued this inhibition, indicating that the blocking effect of the antibodies was specific. Depletion of EEA1 from cytosol caused a

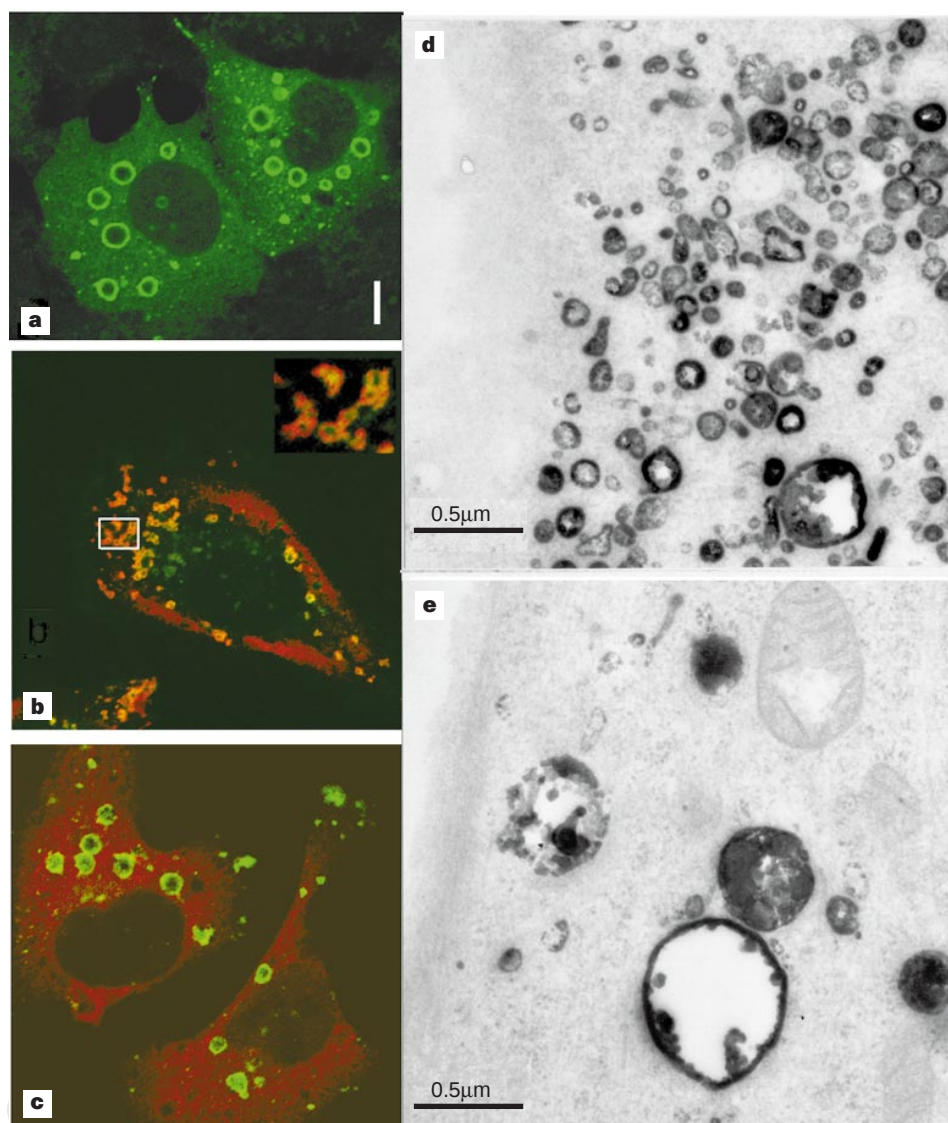


Figure 2 The C-terminal part of EEA1 inhibits basal and Rab5-stimulated early-endosome fusion *in vivo*. **a–c**, Confocal immunofluorescence micrographs showing the intracellular localization of Rab5^{Q79L} (green) and Myc-EEA1-CT (red) in BHK cells transfected with Rab5^{Q79L} (**a**), Rab5^{Q79L} + Myc-EEA1-CT (**b**), or

Rab5^{Q79L} + Myc-EEA1-CT^{C1,358S} (**c**). Yellow indicates colocalization. Scale bar in **a**, 5 μm. **d, e**, Electron micrographs of BHK cells after internalization of horseradish peroxidase for 10 min at 37°C. Myc-EEA1-CT-expressing cells show many small endocytic profiles (**d**), in contrast to early endosomes in control cells (**e**).

somewhat lower inhibition of fusion (50–70%), and this effect could be rescued on addition either of an EEA1-enriched fraction purified from cytosol or of recombinant EEA1 (Fig. 3b). These data indicate that EEA1 is required for homotypic early-endosome fusion.

The observation that EEA1 functions in endosome fusion is intriguing in view of previous findings that PI(3)K activity is required both for endosome fusion^{5–7,17} and for EEA1 binding to endosomes^{11,12}. We found that *in vitro* endosome fusion was inhibited ~70% by 100 nM wortmannin, a PI(3)K inhibitor. This inhibition was counteracted by Rab5 (incorporated as an XTP-binding mutant, Rab5^{D136N}, loaded with XTP-γS or XTP¹⁸), which stimulated fusion strongly (Fig. 3b). The effect of Rab5 (–XTPγS) is probably the result of mass action of activated Rab5 rather than an effect of neutralization of the inhibition by wortmannin⁹. The finding that EEA1 depletion from cytosol only partially inhibited fusion (Fig. 3b) could be explained by the presence of residual EEA1 on the endosome membrane, in agreement with previous localization studies¹⁰. Fusion after EEA1 depletion and wortmannin treatment was less than after either treatment alone, indicating that the two treatments had an additive effect (Fig. 3b). The presence of

Rab5–XTP restored fusion up to only 35% of the basal level under these conditions. Rab5–XTPγS had a similar effect to Rab5–XTP (results not shown). The effect of Rab5–XTP and Rab5–XTPγS is probably due to their action on residual EEA1 present in the fusion reaction. These results show that EEA1 is essential for Rab5-dependent endosome fusion, and that its activity is regulated by PI(3)K.

Given that EEA1 can bind to PtdIns(3)P¹² as well as to activated Rab5, we determined whether Rab5–GTP can counteract the inhibition of early-endosome fusion induced by wortmannin by stabilizing EEA1 on the membrane. For this purpose we used Madin–Darby canine kidney (MDCK) cells expressing Rab5^{Q79L} (Fig. 4)¹⁹. Although EEA1 was abundant in the cytosol (Fig. 4, lane 1) as well as in the membrane fraction (lane 2) of untransfected cells, most EEA1 was membrane-bound in Rab5^{Q79L}-expressing cells (lanes 5, 6). Most important, on treatment with wortmannin, EEA1 efficiently dissociated from the membranes of untransfected cells (lanes 3, 4)^{11,12}, whereas a significant fraction remained membrane-bound in the Rab5^{Q79L}-expressing cells (lanes 7, 8). These data indicate that EEA1 binding to endosomes is dependent on PtdIns(3)P as well as on Rab5–GTP and that the presence of

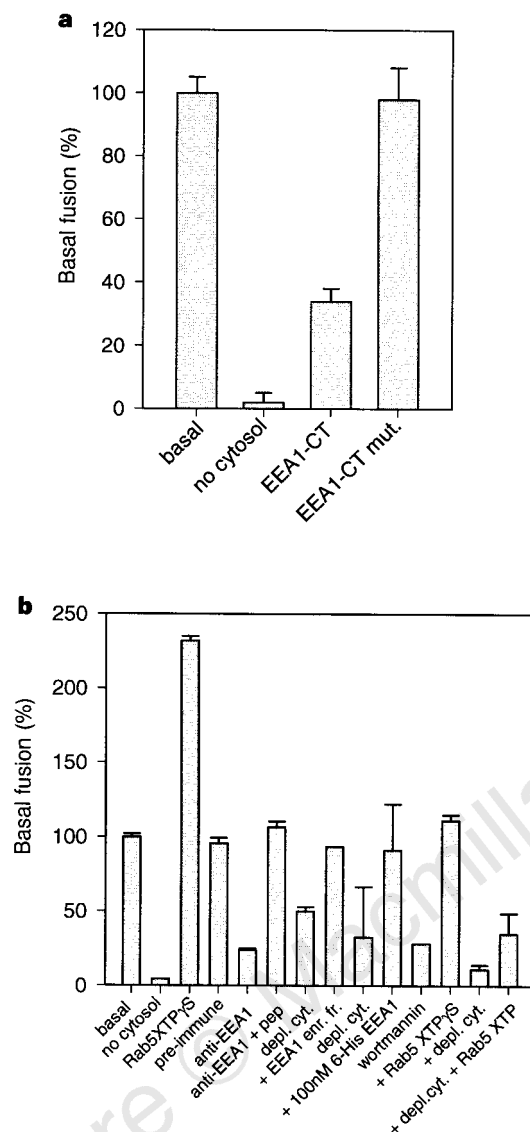


Figure 3 EEA1 is required for early endosome fusion *in vitro*. Homotypic early-endosome fusion in the presence of (a) cytosol from BHK cells transfected with EEA1-CT or EEA1-CT^{C136S} or (b) preimmune serum, anti-EEA1 serum, anti-EEA1 serum + EEA1 N-terminal peptide added, EEA1-depleted HeLa cell cytosol in the absence or presence of recombinant His₆-EEA1, EEA1-enriched cytosolic fraction, or HeLa cell cytosol incubated with wortmannin in the absence or presence of Rab5^{D136N}-XTP-γS or Rab5^{D136N}-XTP (75 nM). Note that the four right-most columns of panel b all represent fusion experiments done in the presence of wortmannin. All fusion activities were calculated as a percentage of that occurring in the presence of cytosol and an ATP-regeneration system (basal fusion). This fusion was typically 5–10% of maximal fusion, defined as the signal detected in the presence of Triton X-100. Fusion in the absence of ATP was less than 10% of basal fusion (data not shown). Error bars represent deviations between duplicates. Note the different scale in a and b.

activated Rab5 on the membrane can compensate for the lack of PtdIns(3)P in the presence of wortmannin.

The findings that EEA1 binds to activated Rab5 and regulates Rab5-dependent membrane fusion establish EEA1 as a new Rab5 effector. The two Rab5 effectors, EEA1 and Rabaptin-5 (ref. 20), are likely to have distinct functions; they are structurally different, and EEA1 is much more abundant than Rabaptin-5 on early endosomes^{10,20}. Because Rab5 is found not only on early endosomes but also at the plasma membrane and on clathrin-coated vesicles^{21,22}, other molecules are needed to confer directionality to Rab5-dependent membrane fusion. The restricted localization

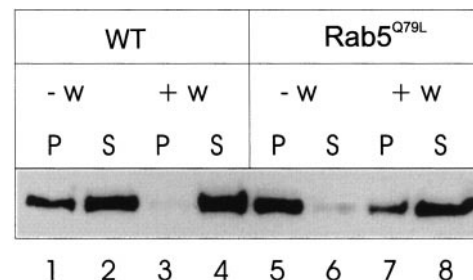


Figure 4 The membrane association of EEA1 is regulated by Rab5 and PI(3)K. Wild-type (lanes 1–4) or Rab5^{Q79L}-transfected¹⁹ (lanes 5–8) MDCK I cells in 9-cm plastic dishes were incubated for 30 min at 37 °C in the absence (lanes 1, 2, 5, 6) or presence (lanes 3, 4, 7, 8) of 100 nM wortmannin. The cells were then homogenized, and postnuclear supernatants were centrifuged at 100,000g for 30 min (ref. 20). The pellet (lanes 1, 3, 5 and 7) and supernatant fractions (lanes 2, 4, 6 and 8) were analysed by SDS-PAGE followed by immunoblotting with anti-EEA1 antibodies.

of EEA1 to early endosomes indicates that it may fulfil such a role.

The following model may explain our results. EEA1 may be recruited initially to early endosomes through binding to Rab5-GTP, and may then be stabilized further on the membrane by interaction of its FYVE finger with PtdIns(3)P. As the affinity of EEA1 for either of these components may be low, or because the components are labile, the binding of EEA1 to membranes would be inefficient in the absence of either Rab5-GTP or PtdIns(3)P. Our model predicts that the presence of either Rab5-GTP or PtdIns(3)P would be sufficient to recruit EEA1 to membranes if present in excessive amounts. This can be seen in the presence of excess Rab5-GTP (Fig. 4) and in experiments using liposomes (with excess PtdIns(3)P)^{11,12}. Our results indicate that one important function of PI(3)K in the endocytic pathway may be to ensure the efficient attachment of the new Rab5 effector EEA1 to membranes. EEA1 is therefore the first molecular link between PI(3)K and Rab5 in endocytic membrane traffic.

Methods

Antibodies and recombinant proteins. Anti-EEA1 was obtained by immunizing a rabbit with a peptide corresponding to the 16 N-terminal residues of EEA1 coupled to diphtheria toxoid. The antiserum was affinity-purified on recombinant MBP-EEA1-NT immobilized on Affi-Gel (Bio-Rad). For expression of GST-Rab5, Rab5 complementary DNA was cloned into pGEX-5x-3 (Pharmacia). In order to express His₆-EEA1 in insect cells, we cloned Myc-EEA1¹³ into pFAST-Bac-HTa (Gibco-BRL) and constructed a recombinant baculovirus according to the manufacturer's instructions. For expression of MBP-EEA1-NT and MBP-EEA1-CT, we cloned the respective EEA1 constructs into pMAL-C2 (New England Biolabs). The *Escherichia coli* expression plasmids were transformed into *E. coli* BL-21(DE3) cells. Expression and protein purification were according to the manufacturers' instructions.

Measurement of protein-protein interactions with the yeast two-hybrid system. Yeast two-hybrid constructs were prepared in the 'bait' vector pLexA/pBTM116 (ref. 23) and the 'prey' vector pGAD-GH (Clontech) using Rab5 (ref. 21) and EEA1 (ref. 10) sequences with the indicated mutations. L40 yeast reporter cells²³ were transformed as indicated in Table 1, and liquid β-galactosidase assays²⁰ were done on duplicate transformants.

Rab5 affinity chromatography. For binding of cytosolic EEA1, MBP-Rab5 was coupled to Affi-Gel 15 (Bio-Rad) and then loaded with GDP or GTP-γS¹⁶. The beads were further incubated with a 0–40% (NH₄)₂SO₄-precipitated fraction of bovine brain cytosol for 1.5 h at 4 °C, and bound proteins were analysed by SDS-PAGE and immunoblotting with anti-EEA1 antibodies. To detect binding of recombinant EEA1, glutathione-agarose beads (Pharmacia) were incubated in the absence or presence of GST-Rab5 (2 μM) for 1 h at 37 °C in a buffer containing 10 μM GDP or GTP-γS. Recombinant His₆-EEA1 (600 nM), MBP-EEA1-NT (1 μM) or MBP-EEA1-CT (43 nM) was added to the beads for 1 h at 4 °C in the presence of 10 mg ml⁻¹ albumin. Bound

proteins were subsequently eluted with 10 mM reduced glutathione for 20 min at room temperature and analysed by SDS-PAGE and immunoblotting with anti-EEA1 or anti-MBP antibodies.

Transient expression in BHK cells. EEA1 and Rab5 constructs were cloned behind the T7 promoter of plasmid pGEM-1 (Promega) and expressed in BHK-21 cells using the T7 RNA polymerase recombinant vaccinia virus system and lipofection as described²⁴. Cells were analysed 4 h after transfection.

Confocal immunofluorescence microscopy. BHK cells grown on coverslips were transiently transfected, fixed and stained for immunofluorescence microscopy as described²⁰. Myc-tagged proteins were detected using the 9E10 mouse monoclonal antibody²⁵, and Rab5^{Q79L} was visualized with affinity-purified anti-Rab5 antibodies²¹, followed by rhodamine- or fluorescein-conjugated secondary antibodies against mouse or rabbit IgG (Jackson ImmunoResearch). Coverslips were examined using a Zeiss confocal scanning laserbeam microscope.

Electron microscopy. BHK cells in 3-cm plastic dishes were incubated for 10 min at 37 °C with horseradish peroxidase (5 mg ml⁻¹, type VI, Sigma), then washed extensively with ice-cold phosphate-buffered saline and fixed with 2% glutaraldehyde in 0.1% sodium cacodylate, pH 7.2. The cells were then processed for electron microscopy²⁶ and 120–250-nm sections were examined in a JEOL 1200 electron microscope.

Assay of early-endosome fusion. Two early-endosome-enriched fractions containing, respectively, biotinylated transferrin or anti-transferrin were separately prepared from HeLa cells, and the fusion was assayed in the presence of HeLa cell cytosol containing an ATP-regenerating system, unless otherwise specified, as described¹⁶. For immunodepletion of EEA1, we used protein-A-Sepharose preincubated with anti-EEA1 antibodies, as described¹ for immunodepletion of Rabaptin-5 (ref. 16). Immunoblotting experiments indicated that >90% of cytosolic EEA1 was depleted under these conditions. The EEA1-enriched fraction used in rescue experiments was obtained by mono-Q ion-exchange chromatography of a 30–40% ammonium sulphate fraction from bovine brain cytosol¹⁶.

Received 1 June; accepted 13 July 1998.

- Novick, P. & Zerial, M. The diversity of Rab proteins in vesicle transport. *Curr. Opin. Cell Biol.* **9**, 496–504 (1997).
- Oikkonen, V. M. & Stenmark, H. Role of Rab GTPase in membrane traffic. *Int. Rev. Cytol.* **176**, 1–85 (1997).
- Roth, M. G. & Sternweis, P. C. The role of lipid signaling in constitutive membrane traffic. *Curr. Opin. Cell Biol.* **9**, 519–526 (1997).
- De Camilli, P., Emr, S. D., McPherson, P. S. & Novick, P. Phosphoinositides as regulators in membrane traffic. *Science* **271**, 1533–1539 (1996).
- Li, G. *et al.* Evidence for phosphatidylinositol 3-kinase as a regulator of endocytosis via activation of Rab5. *Proc. Natl Acad. Sci. USA* **92**, 10207–10211 (1995).
- Jones, A. T. & Clague, M. J. Phosphatidylinositol 3-kinase activity is required for early endosome fusion. *Biochem. J.* **311**, 31–34 (1995).
- Spiro, D. J., Boll, W., Kirchhausen, T. & Wessling-Resnick, M. Wortmannin alters the transferrin receptor endocytic pathway *in vivo* and *in vitro*. *Mol. Biol. Cell* **7**, 355–367 (1996).
- Gorvel, J. P., Chavrier, P., Zerial, M. & Gruenberg, J. Rab5 controls early endosome fusion *in vitro*. *Cell* **64**, 915–925 (1991).
- Jones, A. T., Mills, I. G., Scheidig, A. J., Alexandrov, K. & Clague, M. J. Inhibition of endosome fusion by wortmannin persists in the presence of activated Rab5. *Mol. Biol. Cell* **9**, 323–332 (1998).
- Mu, F. T. *et al.* EEA1, an early endosome-associated protein. EEA1 is a conserved alpha-helical peripheral membrane protein flanked by cysteine “fingers” and contains a calmodulin-binding IQ motif. *J. Biol. Chem.* **270**, 13503–13511 (1995).
- Patki, V., Virbasius, J., Lane, W. S., Toh, B. H., Shpetner, H. S. & Corvera, S. Identification of an early endosomal protein regulated by phosphatidylinositol 3-kinase. *Proc. Natl Acad. Sci. USA* **94**, 7326–7330 (1997).
- Gaullier, J.-M. *et al.* A functional PtdIns(3)P-binding motif. *Nature* **394**, 433–434 (1998).
- Stenmark, H., Aasland, R., Toh, B. H. & D’Arrigo, A. Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger. *J. Biol. Chem.* **271**, 24048–24054 (1996).
- Vitale, G. *et al.* Distinct Rab-binding domains mediate the interaction of rabaptin-5 with GTP-bound Rab4 and Rab5. *EMBO J.* **17**, 1941–1951 (1998).
- Stenmark, H., Parton, R. G., Steele-Mortimer, O., Lütcke, A., Gruenberg, J. & Zerial, M. Inhibition of Rab5 GTPase activity stimulates membrane fusion in endocytosis. *EMBO J.* **13**, 1287–1296 (1994).
- Horiuchi, H. *et al.* A novel Rab5 GTP/GDP exchange factor complexed to rabaptin-5 links nucleotide exchange to effector recruitment and function. *Cell* **90**, 1149–1159 (1997).
- Shpetner, H., Joly, M., Hartley, D. & Corvera, S. Potential sites of PI-3 kinase function in the endocytic pathway revealed by the PI-3 kinase inhibitor, wortmannin. *J. Cell Biol.* **132**, 595–605 (1996).
- Rybin, V. *et al.* GTPase activity of Rab5 acts as a timer for endocytic membrane fusion. *Nature* **383**, 266–269 (1996).
- D’Arrigo, A., Bucci, C., Toh, B. H. & Stenmark, H. Microtubules are involved in bafilomycin A1-induced tubulation and Rab5-dependent vacuolation of early endosomes. *Eur. J. Cell Biol.* **72**, 95–103 (1997).
- Stenmark, H., Vitale, G., Ullrich, O. & Zerial, M. Rabaptin-5 is a direct effector of the small GTPase Rab5 in endocytic membrane fusion. *Cell* **83**, 423–432 (1995).
- Chavrier, P., Parton, R. G., Hauri, H. P., Simons, K. & Zerial, M. Localization of low molecular weight GTP binding proteins to exocytic and endocytic compartments. *Cell* **62**, 317–329 (1990).
- Bucci, C. *et al.* The small GTPase Rab5 functions as a regulatory factor in the early endocytic pathway. *Cell* **70**, 715–728 (1992).
- Vojtek, A. B., Hollenberg, S. M. & Cooper, J. A. Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* **74**, 205–214 (1993).

- Stenmark, H., Bucci, C. & Zerial, M. Expression of Rab GTPases using recombinant vaccinia viruses. *Methods Enzymol.* **257**, 155–164 (1995).
- Evan, G. I., Lewis, G. K., Ramsay, G. & Bishop, J. M. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.* **5**, 3610–3616 (1985).
- Tooze, J. & Hollinshead, M. Tubular early endosome networks in AtT20 and other cells. *J. Cell Biol.* **115**, 635–653 (1991).

Acknowledgements. We thank E. Rønning for technical assistance; D. Warren for help with the baculovirus system; N. Salmon and T. Nordeng for help with confocal microscopy; V. Rybin for providing the Rab5D^{136N}-Rep1 complex; E. Kolpakova for anti-MBP antibodies; and H. McBride, S. Olsnes, K. Sandvig and B. Sønnichsen for comments on the manuscript. H.S. was supported by the Top Research Programme, the Research Council of Norway, the Norwegian Cancer Society and the Novo Nordisk Foundation. This work was supported by a European Community TMR grant (to H.S. and M.Z.).

Correspondence and requests for materials should be addressed to H.S. (e-mail: stenmark@ulrik.uio.no).

Transcriptional activators direct histone acetyltransferase complexes to nucleosomes

Rhea T. Utley*, Keiko Ikeda*†, Patrick A. Grant*, Jacques Côté‡, David J. Steger*, Anton Eberharter*, Sam John* & Jerry L. Workman*

*Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biology and The Center for Gene Regulation, 306 Althouse Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802-4500, USA

†Department of Biology, Jichi Medical School, Minamikawachi, Kawachi, Tochigi 329-04, Japan

‡Laval University Cancer Research Center, Hôtel-Dieu de Québec, Québec G1R-2J6, Canada

Transcriptional co-activators were originally identified as proteins that act as intermediaries between upstream activators and the basal transcription machinery. The discovery that co-activators such as *Tetrahymena* and yeast Gcn5^{1,2}, as well as human p300/CBP^{3,4}, pCAF⁵, Src-1⁶, ACTR⁷ and TAFII250⁸, can acetylate histones suggests that activators may be involved in targeting acetylation activity to promoters. Several histone deacetylases have been linked to transcriptional co-repressor proteins⁹, suggesting that the action of both acetylases and deacetylases is important in the regulation of many genes. Here we demonstrate the binding of two native yeast histone acetyltransferase (HAT) complexes to the herpesvirus VP16 activation domain and the yeast transcriptional activator Gcn4, and show that it is their interaction with the VP16 activation domain that targets Gal4-VP16-bound nucleosomes for acetylation. We find that Gal4-VP16-driven transcription from chromatin templates is stimulated by both HAT complexes in an acetyl CoA-dependent reaction. Our results demonstrate the targeting of native HAT complexes by a transcription-activation domain to nucleosomes in order to activate transcription.

We previously identified four multisubunit native yeast HAT complexes². Two of these complexes, SAGA and Ada, preferentially acetylate nucleosomal histone H3 and contain Gcn5 as the catalytic histone acetyltransferase subunit. SAGA (for Spt-Ada-Gcn5-acetyltransferase) is an 1,800K HAT complex which also contains Ada2, Ada3, Ada5/Spt20, Spt3 and Spt7 (ref. 2) explaining the functional link between the Ada and Spt classes of transcription regulator^{10,11}. The 800K Ada complex contains Ada2 and Ada3 as well as Gcn5, but none of the Spt proteins². Two HATs, formerly termed complex 2, which acetylates histones H4 and H2A, and complex 3, with a substrate preference for histone H3 (ref. 2), were purified and named NuA4 (for nucleosomal acetyltransferase of histone H4) and NuA3, respectively.

Acetylation of histone H3 is stimulated specifically in the promoter regions of Gcn5-dependent genes¹², indicating that promo-

ter-binding factors may recruit HATs to promoter regions. Both Ada2 and Ada3 proteins potentiate transcription of VP16 and Gcn4-dependent reporter genes in yeast^{13,14}. To see whether transcription activators can recruit HAT complexes, we tested for binding of the four native yeast HAT complexes with the VP16 activation domain (Fig. 1a) and Gcn4 (Fig. 1b) by using a glutathione-S-transferase (GST) pull-down assay. After incubating HATs with beads bound to the indicated GST fusion proteins, the supernatants and beads were assayed for their ability to acetylate nucleosomes². Figure 1a, b shows the presence of acetylated histones and therefore of an active HAT complex; control GST alone did not pull down any of the HAT complexes. SAGA and NuA4 activities were depleted in the GST–VP16 and the GST–Gcn4 supernatants and were recovered on the beads. NuA3 did not bind GST–VP16 or Gcn4. The Ada complex, which contains at least three of the same subunits as SAGA, did not interact with the VP16 activation domain or with Gcn4, although Ada2 does^{15,16}; this suggests that Ada2 may be masked in the native Ada complex. As the yeast activator Gcn4 interacts with SAGA but not with Ada, recruitment of the SAGA complex to the promoter of Gcn5-dependent genes is probably necessary for their expression.

Deletion of the carboxy-terminal amino-acid residues 456–490 of VP16 results in a 30–50% reduction in transcriptional activity^{17,18}, and a phenylalanine-to-proline (F442P) mutation in this truncated protein prevents transcription^{18–20}. Interaction of VP16 with several basal factors is also disrupted by the F442P mutation^{21–23}. To assess the specificity of SAGA and NuA4 binding with VP16, we used mutant forms of the VP16 activation domain (Fig. 2a) in GST pull-down assays. Compared with full-length VP16 (Fig. 2b, lanes 3, 7), the amino-terminal domain interacted less strongly with SAGA (lane 8), which may be indicative of a weaker interaction of VP16(Δ 456) with Ada2 (refs 15, 16). Although the GST–VP16(F442P) is as acidic as GST–VP16(Δ 456), it did not bind the SAGA complex (Fig. 2b, lane 9). GST pull-down assay using purified NuA4 (Fig. 2c) gave similar results, so the interaction of SAGA and NuA4 with the VP16 activation domain is not just

ionic. Rather, the apparent strength of the SAGA and NuA4 interaction with different VP16 domains reflects the relative strengths of the domains as transcription activators *in vivo*^{18–20}.

The fact that both SAGA and NuA4 interact with VP16 and Gcn4 indicates that these HATs could be targeted to specific nucleosomes as a result of this binding. We therefore designed an assay to determine the ability of Gal4–VP16 to recruit HAT complexes for acetylating factor-bound nucleosomes by using reconstituted nucleosome core particles containing 183-base-pair (bp) DNA fragments with a single Gal4-binding site (Fig. 3a). The Gal4 site lies only 20 bp from the end of this fragment, so the binding of Gal4 derivatives is less restricted by the presence of the nucleosome²⁴ (data not shown). We added increasing amounts of Gal4–VP16 or Gal4(1–94), which contains only the DNA-binding and dimerization domains of Gal4, to nucleosomes. At low concentrations of Gal4 derivatives, only a small fraction of total nucleosomes bind (Fig. 3b, top, lanes 2 and 3). Fluorography of ³H-labelled acetate incorporated by the SAGA complex revealed that this small fraction of nucleosomes bound by Gal4–VP16 was a preferred target for acetylation by the SAGA complex (Fig. 3b, bottom, lanes 2 and 3). The VP16 activation domain was required for this preferential acetylation because it did not occur with the DNA-binding domain of Gal4 alone (Fig. 3b: compare top and bottom panels, lanes 5–7). Likewise, Gal4–VP16 could also direct nucleosome acetylation by the NuA4 complex (Fig. 3c). Thus, although the SAGA and NuA4 complexes could acetylate nucleosomes in the absence of activator, prebinding of Gal4–VP16 recruited these HAT complexes to nucleosomes and caused increased acetylation of factor-bound nucleosome cores. We also tested the Ada and NuA3 complexes in the targeting assay (Fig. 3d, e). In agreement

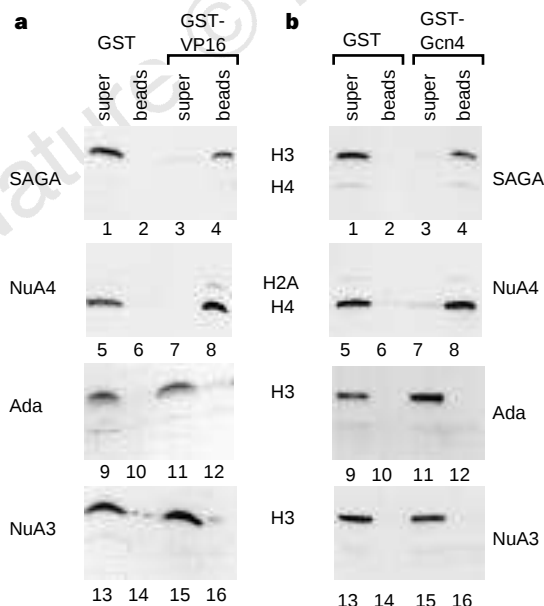


Figure 1 Two distinct HAT complexes interact with the VP16 activation domain and the yeast activator Gcn4. **a**, GST pull-down assays were done with either GST–VP16 or GST alone bound to glutathione–Sepharose beads and the indicated HAT activity. Supernatants and beads were assayed for nucleosomal HAT activity and reactions were separated on SDS–PAGE. Acetylation of histones indicates the presence of SAGA (lanes 1–4), NuA4 (lanes 5–8), Ada (lanes 9–12), or NuA3 (lanes 13–16). **b**, Assays were done as in **a**, except that the interaction with GST–Gcn4 was tested.

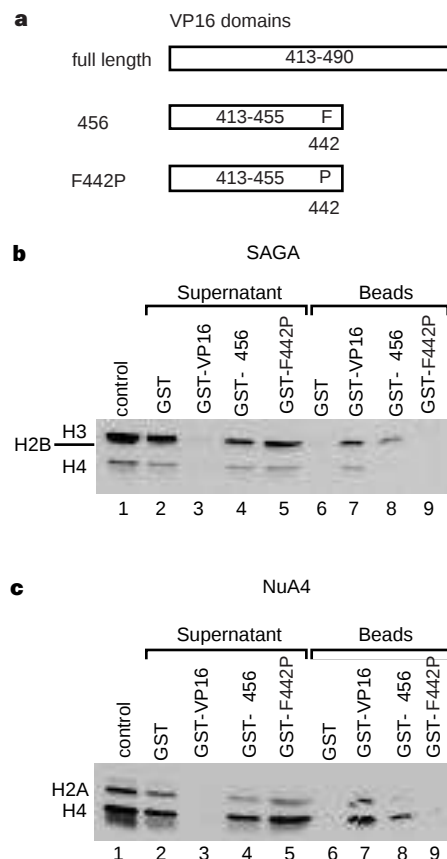


Figure 2 SAGA and NuA4 interact specifically with the VP16 activation domain. **a**, VP16 activation domains and mutants used in the GST pull-down assay. **b**, **c**, SDS–PAGE fluorogram of nucleosomes acetylated by the SAGA complex (**b**) or the NuA4 complex (**c**).

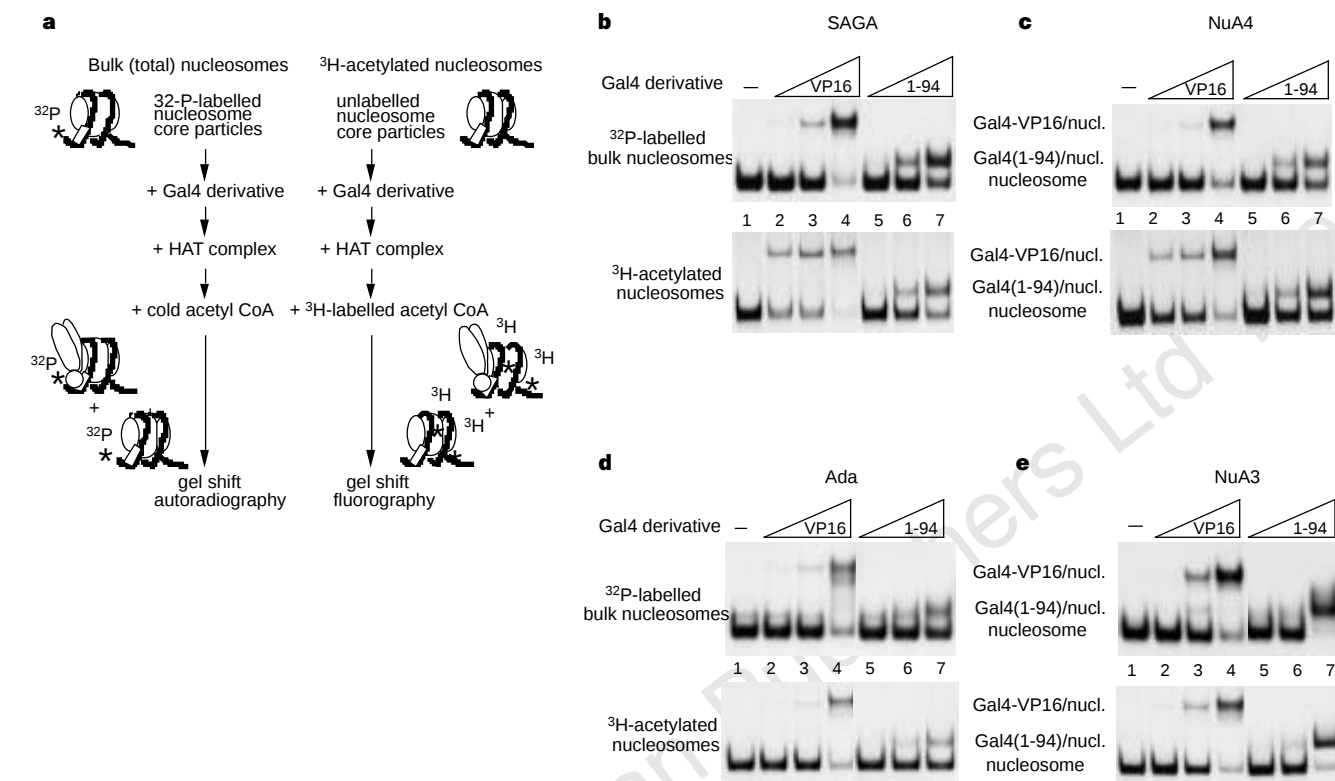


Figure 3 VP16 directs acetylation by SAGA and NuA4. **a**, The targeting assay. **b**, Comparison of bulk nucleosomes to nucleosomes acetylated by SAGA in the absence (lane 1) and presence of 4, 12 and 40 nM (lanes 2–4) Gal4–VP16 or 24, 80 and 240 nM (lanes 5–7) Gal4(1–94). SAGA complex was present in all reactions. **c**,

Targeting of NuA4 by VP16. Gal4 concentrations were as in **b**. **d**, Ada was not targeted by VP16. Gal4 concentrations were as in **b**. **e**, VP16 does not target NuA3. Lanes 1–4 contained 0, 8, 24 and 80 nM Gal4–VP16. Gal4(1–94) was added at 40, 120 and 400 nM in lanes 5–7.

with the GST pull-down assays, the binding of Gal4–VP16 did not cause increased acetylation by Ada or NuA3 of bound nucleosomes.

As only HATS that interacted with VP16 could increase acetylation of bound nucleosomes, this increase is not a result of acetylation favouring binding of the activator. To verify this, we did control experiments in which nucleosomes were preacetylated with either SAGA or NuA4 which were then heat-inactivated (data not shown) before Gal4 derivatives were added. The binding of Gal4–VP16 to preacetylated nucleosomes was not stimulated (Fig. 4a, b). Thus, these HAT complexes were targeted to nucleosomes prebound by activators (Fig. 3b, c).

We tested the function of these two HAT complexes by assaying SAGA and NuA4 for their ability to stimulate VP16-driven transcription from nucleosome arrays. By placing a minimal E4 promoter containing five Gal4 sites within a spaced array of nucleosome-positioning sequences on 5S ribosomal DNA, we constructed a template in which two nucleosomes were formed over the promoter region in phase with a repeated 5S nucleosome array (Fig. 5a); their position was confirmed by mapping with micrococcal nuclease (data not shown). Nucleosomal templates were preincubated with the indicated Gal4 derivatives and HAT complexes, and transcription was assayed. Transcription was almost undetectable from these templates in the absence of activator or in the presence of Gal4 DNA-binding domain alone (Fig. 5b: lanes 1, 2, 4, 5), indicating that the VP16 activation domain is required for significant transcription from these nucleosomal templates. Gal4–VP16-driven transcription from these templates was stimulated by the SAGA (Fig. 5b, lane 6) and NuA4 complexes (Fig. 5b, lane 12), but we were unable to detect an increase in transcription in the presence of SAGA or NuA4 and Gal4(1–94) (Fig. 5b, lanes 5, 11); neither complex stimulated transcription from naked DNA (Fig. 5c). This HAT-dependent stimulation also required acetyl CoA (Fig. 5d: compare lanes 5 to

6 and 11 to 12). Thus, stimulation of transcription by SAGA and NuA4 depends on their acetyltransferase activity and occurs only on nucleosomal substrates, indicating that histone acetylation has a role in transcriptional activation.

Our results reveal a specific functional binding between the yeast activator Gcn4 or the VP16 activation domain with the SAGA and NuA4 HAT complexes. We have shown that two other previously identified yeast HAT complexes, Ada and NuA3, do not interact with VP16 or Gcn4, indicating that the interaction of activator and HAT must be specific. We have also demonstrated that a transcriptional activator can direct nucleosomal acetylation by recruiting HAT complexes in order to stimulate activated transcription. □

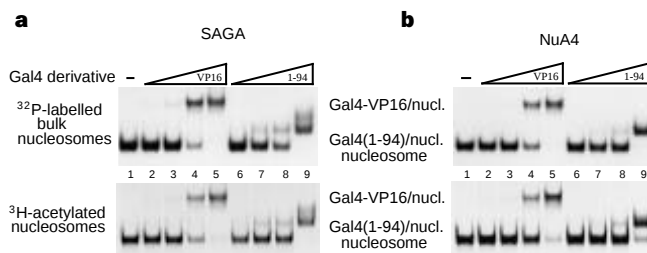


Figure 4 Preacetylation by SAGA and NuA4 precludes targeting by the VP16 activation domain. Reactions were done under conditions similar to the targeting assay except that nucleosome acetylation was for 30 min at 30°C with HAT complex and unlabelled or ^3H -labelled acetyl CoA. HAT complexes were inactivated, and binding of Gal4 derivatives was for 30 min at 30°C. Samples contained 0, 4, 12, 40 and 120 nM Gal4–VP16 (lanes 1–5) or 12, 40, 120 and 400 nM Gal4(1–94) (lanes 6–9). **a**, Reactions with SAGA present in all lanes; **b**, NuA4 present in all lanes.

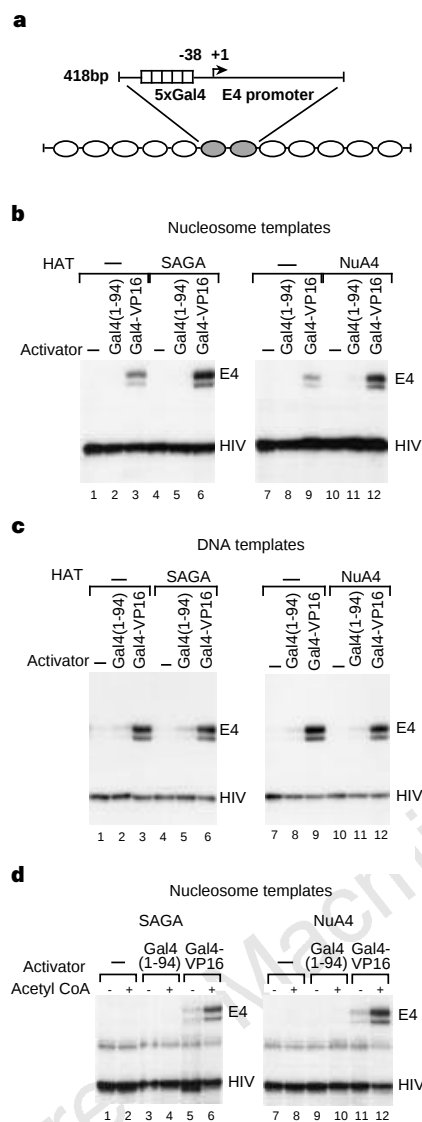


Figure 5 SAGA and NuA4 stimulate Gal4-VP16-dependent transcription. **a**, The nucleosomal G5E4 5S transcription template. **b**, The nucleosomal template was transcribed following binding with Gal(1-94), Gal-VP16, or no activator in the presence of SAGA or NuA4, as indicated. The 81-base transcript of HIV served as a recovery control. **c**, No stimulation of VP16-dependent transcription by SAGA or NuA4 was observed on DNA templates. **d**, Transcription was assayed with nucleosomal templates as in **a**, including or omitting acetyl CoA as indicated. SAGA or NuA4 stimulation of VP16-mediated transcription required acetyl CoA.

Methods

HAT purification. The preparation of yeast extracts and separation of the SAGA, Ada, NuA4 and NuA3 complexes was done as described², except that the order of columns was modified. SAGA, Ada or NuA4 were purified over Ni²⁺-NTA agarose (Qiagen), followed by Mono-Q HR 5/5 (Pharmacia), Mono S HR 5/5 (Pharmacia), histone agarose (Sigma), and Superose-6 HR 10/30 (Pharmacia). NuA3 was purified similarly, except that a heparin-Sepharose column was substituted for Superose-6. HAT complexes were purified ~70,000–100,000-fold from yeast whole-cell extracts. Typically, we used 1 µl peak activity from the final column for the targeting and transcription assays: 1 µl of these fractions generated between 3,000 and 6,000 c.p.m. in our standard nucleosomal HAT assay². Western blots detected no SWI/SNF complex in any of the final HAT fractions.

Protein expression and purification. GST-VP16, GST-Δ456, GST-FP442, GST²² and GST-GCN4 (ref. 16) were expressed in bacteria and purified as described¹⁶. Proteins were quantified by scanning Coomassie-blue stained

SDS-PAGE minigels. Gal4(1-94) was purified as described²⁵, as was Gal4-VP16 (ref. 26), except that the protein was step-eluted at 0.4 M NaCl for the DEAE column. Protein concentrations were determined by the Bradford method²⁷.

GST pull-down assay. Each HAT complex was pre-cleared with glutathione-Sepharose beads and the equivalent of ~2 µl (6,000–12,000 c.p.m.) of HAT complex was incubated with the indicated GST fusion protein for 2 h at 4 °C, with mixing. Supernatant was removed, beads were washed 4 times in bead-wash buffer (150 mM NaCl, 25 mM HEPES, pH 7.5, 5 mM Tris, 1 mM EDTA, 0.05% NP-40, 5 mM DTT, 0.5 mM PMSF, 10% glycerol, 50 µg ml⁻¹ BSA), and equal fractions of both supernatants and beads were assayed for nucleosomal acetyltransferase activity as described². The final assay conditions were adjusted to 50 mM Tris-HCl, pH 8, 10 mM sodium butyrate, 0.5 µg oligonucleosomes, 0.125 µCi ³H-acetyl CoA.

Plasmids and probes. The single Gal4-site probe used for Fig. 2 was generated by PCR using either unlabelled or ³²P-kinased GUB 5' primer from pGALUSFBend as described²⁸. To generate plasmid pIC208G5E4 for transcription, a fragment was amplified from pG5E4T²⁵ using the following primers containing *Xho*I sites: G5E4, 5'-CCCGCTCGAGTGCATGCCTGCAGGTC and G5E4, 3'-CCCGCTCGAGATTACAGCCCCXATAGG. The PCR product was digested with *Xho*I and inserted into *Xho*I-digested pIC208S5²⁹. The 12-nucleosome-long fragment was obtained by digestion of pIC208G5E4 with *Asp*718 and *Cl*at.

Histone preparation and nucleosome reconstitution. Core histones and oligonucleosomes from HeLa cells were prepared as described³⁰. Nucleosomes were reconstituted by octamer transfer as described^{30,31}; core histone was reconstituted as before²⁹.

Targeting assay. 15 µl unlabelled or ³²P-labelled nucleosome cores (150 ng DNA) were incubated with 10 mM sodium butyrate and 20 mM HEPES, pH 7.5, in 30 µl, and HAT activity was assayed in reconstitution buffers^{30,31}. Final salt conditions ranged from 60–75 mM NaCl. Samples were incubated with the indicated Gal4 derivatives for 10 min at 30 °C, followed by a 10-min incubation in the presence of SAGA, NuA4, Ada or NuA3. Finally, 1.7 µM unlabelled or 0.25 µCi ³H-labelled acetyl CoA was added and reacted for 30 min at 30 °C. For the control experiments shown in Fig. 4, all reagents except the Gal4 derivatives were incubated for 30 min at 30 °C for acetylation of nucleosomes. Reactions were then heated to 50 °C for 12 min to inactivate HATs. Gal4 derivatives were added and incubated for an additional 30 min at 30 °C. Samples were run for 3 h on 6% polyacrylamide (29:1, acrylamide:bisacrylamide) in 1 × TBE. ³²P-labelled gels were dried and autoradiographed. ³H-labelled gels were fixed with 40% methanol, 10% acetic acid solution and treated with ENHANCE (Dupont/NEN) for fluorography, according to the manufacturer's instructions.

In vitro transcription. Transcription reactions were essentially carried out as described²⁹, except that 15–21 ng of reconstituted G5E4 dinucleosome array (Fig. 5b, d) or naked DNA (Fig. 5c) were assayed with the indicated amounts of Gal4-VP16 or Gal4(1-94) in the absence or presence of HAT complex for 1–20 min; 0.4 ng HIV plasmid was added with the transcription mix as a recovery control. Primers: 10–20 fmol E4 primer (+86 to +100) or 2–5 fmol HIV primer (+50 to +81).

Received 2 April; accepted 5 June 1998.

- Brownell, J. E. & Allis, C. D. Special HATs for special occasions: Linking histone acetylation to chromatin assembly and gene activation. *Curr. Opin. Genet. Dev.* **6**, 176–184 (1996).
- Grant, P. A. *et al.* Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev.* **11**, 1640–1650 (1997).
- Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953–959 (1996).
- Bannister, A. J. & Kouzarides, T. The CBP coactivator is a histone acetyltransferase. *Nature* **384**, 641–643 (1996).
- Yang, X.-J., Ogryzko, V., Nishikawa, J., Howard, B. H. & Nakatani, Y. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* **382**, 319–324 (1996).
- Spencer, T. E. *et al.* Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature* **389**, 194–198 (1997).
- Chen, H. *et al.* Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell* **90**, 569–580 (1997).
- Mizzen, C. A. *et al.* The TAFII250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**, 1261–1270 (1996).
- Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12**, 599–606 (1998).
- Marcus, G. A., Horieuchi, J., Silverman, N. & Guarente, L. ADA5/SPT20 links the ADA and SPT genes, which are involved in yeast transcription. *Mol. Cell. Biol.* **16**, 3197–3205 (1996).
- Roberts, S. M. & Winston, F. SPT20/ADA5 encodes a novel protein functionally related to the TAT-binding protein and important for transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 3206–3213 (1996).

12. Kuo, M.-H., Zhou, J., Jambeck, P., Churchill, M. E. A. & Allis, C. D. Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes *in vivo*. *Genes Dev.* **12**, 627–639 (1998).
13. Berger, S. L. *et al.* Genetic isolation of ADA2: a potential transcriptional adaptor required for function of certain acidic activation domains. *Cell* **70**, 251–265 (1992).
14. Piña, B. *et al.* ADA3: a gene, identified by resistance to GAL4-VP16, with properties similar to and different from those of ADA2. *Mol. Cell. Biol.* **13**, 5981–5989 (1993).
15. Silverman, N., Agapite, J. & Guarente, L. Yeast ADA2 protein binds to the VP16 protein activation domain and activates transcription. *Proc. Natl Acad. Sci. USA* **91**, 11665–11668 (1994).
16. Barlev, N. A. *et al.* Characterization of physical interactions of the putative transcriptional adaptor, ADA2, with acidic activation domains and TATA-binding protein. *J. Biol. Chem.* **270**, 19337–19344 (1995).
17. Triezenberg, S. J., Kingsbury, R. C. & McKnight, S. L. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* **2**, 718–729 (1988).
18. Berger, S. L., Cress, W. D., Cress, A., Triezenberg, S. J. & Guarente, L. Selective inhibition of activated but not basal transcription by the acidic activation domain of VP16: Evidence for transcriptional adaptors. *Cell* **61**, 1199–1208 (1990).
19. Regier, J. L., Shen, F. & Triezenberg, S. J. Pattern of aromatic and hydrophobic amino acids critical for one of two subdomains of the VP16 transcription activator. *Proc. Natl Acad. Sci. USA* **90**, 883–887 (1993).
20. Cress, W. D. & Triezenberg, S. J. Critical structural elements of the VP16 transcriptional activation domain. *Science* **251**, 87–90 (1991).
21. Ingles, C. J., Shales, M., Cress, W. D., Triezenberg, S. J. & Greenblatt, J. Reduced binding of TFIID to transcriptionally compromised mutants of VP16. *Nature* **351**, 588–590 (1991).
22. Lin, Y.-S. & Green, M. G. Mechanism of action of an acidic transcriptional activator *in vitro*. *Cell* **64**, 971–981 (1991).
23. Xiao, H. *et al.* Binding of basal transcription factor TFIID to the acidic activation domains of VP16 and p53. *Mol. Cell. Biol.* **14**, 7013–7024 (1994).
24. Vettese-Dadey, M., Walter, P., Chen, H., Juan, L.-J. & Workman, J. L. Role of the histone amino termini in facilitated binding of a transcription factor, GAL4-AH, to nucleosome cores. *Mol. Cell. Biol.* **14**, 970–981 (1994).
25. Lin, Y.-S., Carey, M. F., Ptashne, M. & Green, M. R. GAL4 derivatives function alone and synergistically with mammalian activators *in vitro*. *Cell* **54**, 659–664 (1988).
26. Chasman, D. I., Leatherwood, J., Carey, M., Ptashne, M. & Kornberg, R. D. Activation of yeast polymerase II transcription by herpes virus VP16 and GAL4 derivatives *in vitro*. *Mol. Cell. Biol.* **9**, 4746–4749 (1989).
27. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analyt. Biochem.* **72**, 248–254 (1976).
28. Juan, L.-J., Utley, R. T., Vignali, M., Bohm, L. & Workman, J. L. H1-mediated repression of transcription factor binding to a stably positioned nucleosome. *J. Biol. Chem.* **272**, 3635–3640 (1997).
29. Steger, D. J., Eberharter, A., John, S., Grant, P. A. & Workman, J. L. Purified histone acetyltransferases stimulate HIV-1 transcription from preassembled nucleosomal arrays. *Proc. Natl Acad. Sci. USA* (submitted).
30. Côté, J., Utley, R. T. & Workman, J. L. Basic analysis of transcription factor binding to nucleosomes. *Meth. Mol. Genet.* **6**, 108–129 (1995).
31. Utley, R. T. *et al.* *In vitro* analysis of transcription factor binding to nucleosomes and nucleosome disruption/displacement. *Meth. Enzymol.* **274**, 276–291 (1996).

Acknowledgements. We thank M. Green and J. Reese for the gift of GST-VP16 and GST-VP16 mutant expression plasmids, N. Barlev and S. Berger for the GST-GCN4 expression plasmid, and members of J.L.W.'s laboratory for comments on the manuscript. This work was supported by a grant from the National Institute of General Medical Sciences. P.G. is supported by a postdoctoral fellowship from the American Cancer Society; J. C. was a recipient of an MRC Centennial fellowship; A.E. holds a fellowship from the Austrian Science Foundation; D.J.S. was supported by a postdoctoral fellowship from the Cancer Research Institute; S.J. is an HHMI postdoctoral associate and J.L.W. was a Leukemia Society scholar and is currently an HHMI investigator.

Correspondence and requests for materials should be addressed to J.L.W. (e-mail: jlw10@psu.edu).

Structure of the metal-ion-activated diphtheria toxin repressor/tox operator complex

André White*, Xiaochun Ding*†, Johanna C. vanderSpek‡, John R. Murphy‡ & Dagmar Ringe*

* Rosenstiel Basic Medical Sciences Research Center MS029, and
 † Departments of Biochemistry and Chemistry, Brandeis University, Waltham,
 Massachusetts 02454-9110, USA

‡ Department of Medicine, Boston University School of Medicine, Boston,
 Massachusetts 02118-2393, USA

The virulent phenotype of the pathogenic bacterium *Corynebacterium diphtheriae* is conferred by diphtheria toxin, whose expression is an adaptive response to low concentrations of iron. The expression of the toxin gene (*tox*) is regulated by the repressor DtxR (ref. 1), which is activated by transition metal ions. X-ray crystal structures of DtxR with^{2–5} and without (apo-form²) its coordinated transition metal ion have established the

general architecture of the repressor, identified the location of the metal-binding sites, and revealed a metal-ion-triggered subunit–subunit ‘caliper-like’ conformational change. Here we report the three-dimensional crystal structure of the complex between a biologically active Ni(II)-bound DtxR(C102D) mutant, in which a cysteine is replaced by an aspartate at residue 102, and a 33-base-pair DNA segment containing the toxin operator *toxO*. This structure shows that DNA interacts with two dimeric repressor proteins bound to opposite sides of the *tox* operator. We propose that a metal-ion-induced helix-to-coil structural transition in the amino-terminal region of the protein is partly responsible for the unique mode of repressor activation by transition metal ions.

The three-dimensional crystal structure of the active diphtheria toxin repressor mutant DtxR(C102D) in complex with both co-repressor Ni(II) ions and a synthetic oligonucleotide, bearing the corynebacteriophage β *tox* operator sequence (Fig. 1), has been determined by molecular replacement at 3 Å resolution. The stoichiometry of protein to DNA is surprising: two dimers of Ni(II)-DtxR(C102D) interact with the *tox* operator (Fig. 2). The two dimers bind DNA at sites five nucleotides apart, to symmetrically disposed regions of the major groove on almost opposite sides of the centre of the interrupted palindromic operator sequence. A non-crystallographic two-fold axis intersecting the nucleic acid at the central G-C base pair relates these two dimers; however, the two protein dimers do not interact with each other. The stoichiometry is similar in the otherwise completely different *met* repressor⁶.

DNAse I protection assays have shown a footprint for Ni(II)-activated DtxR with the *tox* operator that spans 31 base pairs (bp) on the coding strand and 26 bp on the non-coding strand⁷. It has been unclear how a short helix–turn–helix segment (residues 27–50 of DtxR) in a 226-residue repressor could account for such a large footprint: one explanation was that the carboxy-terminal domain of the repressor, which is partially disordered in the crystal structures of the apo- and holo-repressors^{2–5}, might adopt a conformation that enables it to interact with the operator sequence. Each Ni(II)-DtxR(C102D) dimer interacts with a nucleic acid region of 19 bp; together, both dimers bind to 24 bp and cover approximately 27 bp of DNA (Fig. 2a). This observation, combined with the unique configuration of the two dimers interacting with two major groove regions on almost opposite sides of the double helix, explains the

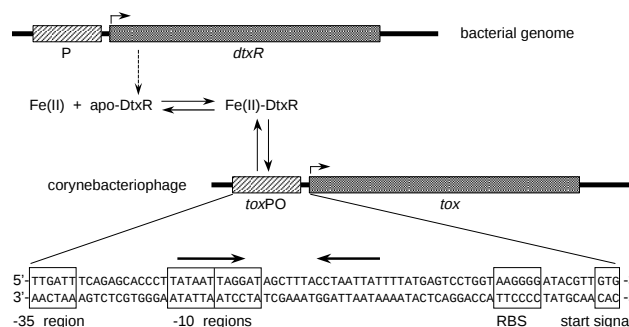


Figure 1 The *trans*-regulated expression of the diphtheria toxin gene, *tox*, in *C. diphtheriae*. The *tox* gene is located on the genome of a corynebacteriophage infecting the bacterium, and its expression is negatively controlled by the iron-activated diphtheria *tox* repressor (DtxR)²⁷. When the level of iron is high, the Fe(II)-DtxR complex binds specifically to the *tox* promoter-operator, *toxPO*, sequence, preventing transcription. When iron becomes the growth-rate limiting nutrient, the ternary complex, Fe(II)-DtxR-*tox* operator, dissociates allowing transcription to proceed. The nucleic acid sequences of the *toxPO* regulatory region are highlighted: –35 and –10 regions of the promoter, the 27-bp interrupted palindromic sequence of the *tox* operator (bold arrows), the ribosome binding site (RBS), and *tox* translational initiation (start signal) are shown.

† Present address: Wadsworth Center, room C160, PO Box 509, Albany, New York 12201-0509, USA.

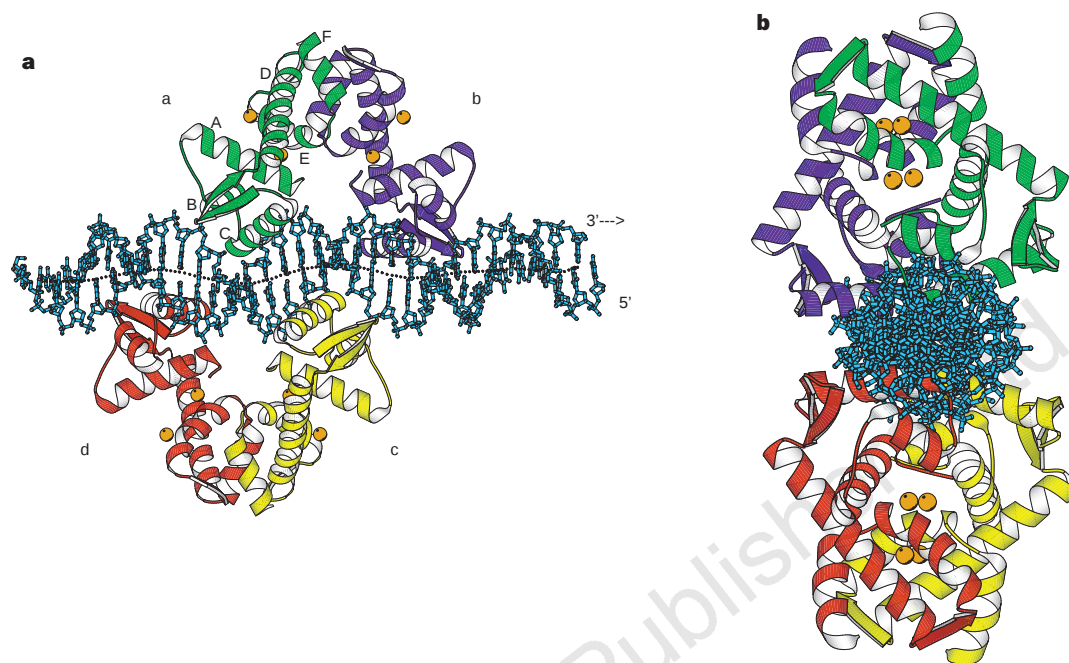


Figure 2 Structure of the Ni(II)-DtxR(C102D)-*tox* operator complex. **a, b**, In these orthogonal views, residues 3–120 of DtxR(C102D) subunits (labelled a to d) are shown, in which ribbons and arrows illustrate α -helices (labelled A–F) and β -strands, respectively, and orange balls represent nickel atoms. The interrupted palindromic 33-base-pair DNA segment (cyan) that includes the *tox*-operator

sites on a molecular pseudo-twofold symmetry axis. Dots are drawn to show the helical axis of the nucleic acid segment, as analysed with CURVES52²⁸; in canonical DNA, this axis would be linear. One dimer (subunits a and b) binds closer to the 3' end of the coding strand (3'→), the other (subunits c and d) closer to the 5' end. These pictures were prepared with MOLSCRIPT²⁹.

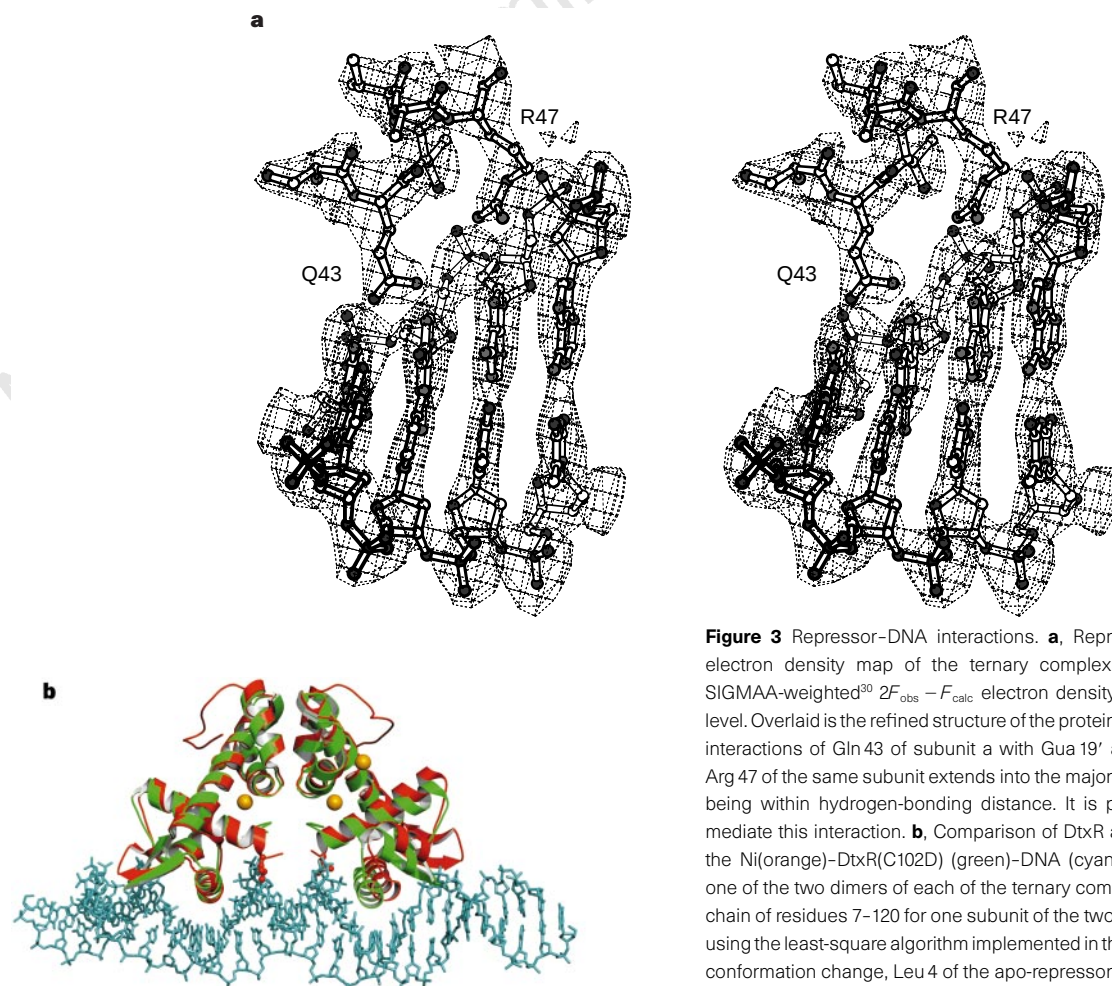


Figure 3 Repressor-DNA interactions. **a**, Representative region of the final electron density map of the ternary complex Ni(II)-DtxR(C102D)-DNA. This SIGMAA-weighted³⁰ $2F_{\text{obs}} - F_{\text{calc}}$ electron density map is contoured at the 1.6σ level. Overlaid is the refined structure of the protein-DNA complex. Shown are the interactions of Gln 43 of subunit a with Gua 19' and Cyt 20'. The side chain of Arg 47 of the same subunit extends into the major groove toward Ade 18' without being within hydrogen-bonding distance. It is possible that water molecules mediate this interaction. **b**, Comparison of DtxR as the apo-repressor (red) and the Ni(orange)-DtxR(C102D) (green)-DNA (cyan) ternary complex. For clarity, one of the two dimers of each of the ternary complexes is not shown. The main chain of residues 7–120 for one subunit of the two structures was superimposed using the least-square algorithm implemented in the program O (ref. 23). Without a conformation change, Leu 4 of the apo-repressor would clash with a phosphate group, as shown.

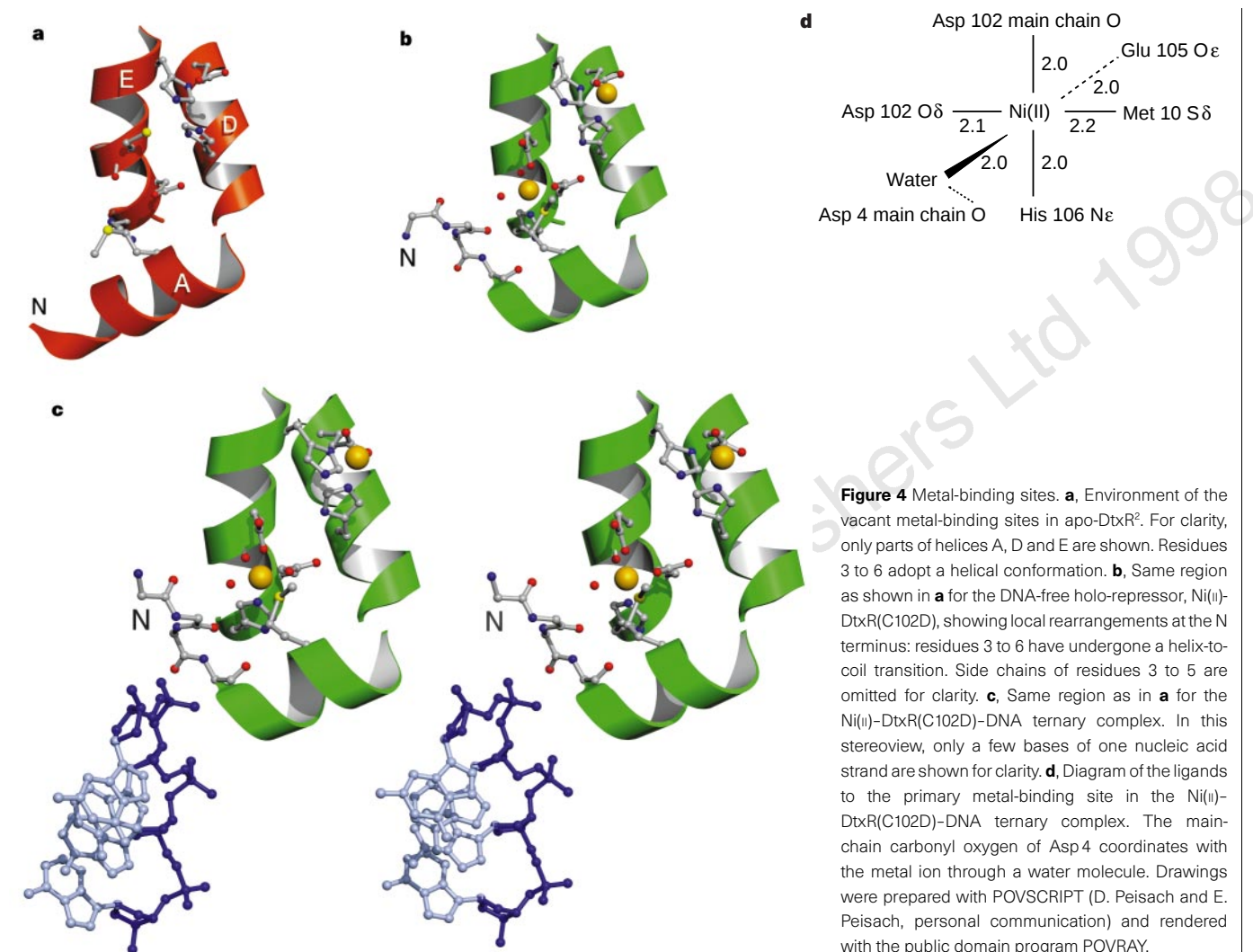


Figure 4 Metal-binding sites. **a**, Environment of the vacant metal-binding sites in apo-DtxR². For clarity, only parts of helices A, D and E are shown. Residues 3 to 6 adopt a helical conformation. **b**, Same region as shown in **a** for the DNA-free holo-repressor, Ni(II)-DtxR(C102D), showing local rearrangements at the N terminus: residues 3 to 6 have undergone a helix-to-coil transition. Side chains of residues 3 to 5 are omitted for clarity. **c**, Same region as in **a** for the Ni(II)-DtxR(C102D)-DNA ternary complex. In this stereoview, only a few bases of one nucleic acid strand are shown for clarity. **d**, Diagram of the ligands to the primary metal-binding site in the Ni(II)-DtxR(C102D)-DNA ternary complex. The main-chain carbonyl oxygen of Asp4 coordinates with the metal ion through a water molecule. Drawings were prepared with POVSCRIPT (D. Peisach and E. Peisach, personal communication) and rendered with the public domain program POVRAY.

biochemical data without invoking involvement of the C-terminal domain. It is also consistent with recent studies of mutants indicating that the C-terminal domain of DtxR is not essential for repressor activity⁸. Although the structure of the ternary complex has been solved with the active C102D mutant of DtxR, we expect the wild-type repressor to behave similarly, based on biochemical and genetic evidence⁹.

Analysis of the ternary complex between the Ni(II)-activated DtxR(C102D) repressor and its cognate DNA reveals a helix-turn-helix (HTH) motif (residues 27–50; helices B and C) that binds to the major groove of the nucleic acid (Fig. 3b). From this motif, residues Arg 27, Ala 28 and Arg 29 of α -helix B, Glu 36 and Ser 37 located on the loop following α -helix B, Thr 40, Ser 32, Arg 47 and Arg 50 of α -helix C interact with phosphate groups of the oligonucleotide. Several residues of this motif change their orientation so that their side chains interact with DNA. The loop of residues 57–61, which joins β -strands 1 and 2 of an antiparallel β -sheet, rearranges so that the guanidium group of Arg 60 is now inserted into the minor groove of the nucleic acid. Thr 7 also interacts with a phosphate group as a consequence of N-terminal rearrangement of the protein. Only Gln 43 interacts directly with the nucleic acid bases (Fig. 3a); other reports have also indicated that most protein–DNA interactions involve nucleic acid phosphate groups rather than bases^{10,11}. The importance of the HTH motif in DtxR is supported by mutational analysis. Deletion of the amino-terminal 47 residues, corresponding to most of the HTH region, and individual mutations at Thr 40, Thr 44, Arg 47 and

Arg 50, each cause the repressor to be inactive, indicating that the HTH region is essential for nucleic acid binding^{2,7,12–14}.

Several transition metal ions, such as Fe(II), Ni(II), Co(II), Cd(II), Mn(II) and, to some extent, Zn(II), activate apo-DtxR and thus confer DNA-binding ability *in vitro*^{3,7,9,12}. The structure of an active mutant DtxR(C102D) and site-directed mutagenesis of DtxR indicate that, of the two metal-ion-binding sites identified earlier, only one is essential for repressor activity^{3,7,9} and that metal-ion binding is cooperative¹⁴. As the protein binds metal ions in the absence of DNA⁷, we expected that the metal ion does not form a bridge between protein and nucleic acid. In fact, our structure shows that the metal ion activates repressor binding to DNA by altering the structure of the protein to allow recognition. Upon binding of Ni(II) to DtxR(C102D), there are two structural changes.

First, in the holo-repressor dimer a slight closure of the N-terminal domains (residues 3–120) occurs upon binding of metal ion, so that α -helices C move closer together by 2 Å (ref. 2). This caliper-like movement was proposed to bring these helices into better alignment with the major grooves of the cognate B-form DNA. Comparison of the apo-repressor structure with that of the ternary complex confirms that this caliper-like conformational change occurs and shows that DNA binding does not lead to a further large displacement (Fig. 3b). Second, upon activation with the co-repressor Ni(II), the six amino-terminal residues of DtxR(C102D) undergo a helix-to-coil transition (Fig. 4). This conformational change is metal-ion-triggered rather than DNA-induced, because the holo-repressor structure of Ni(II)-

DtxR(C102D) in the absence of nucleic acid shows the same rearrangement. This rearrangement of about one helical turn of α -helix A relieves what would otherwise, in the apo-repressor conformation, be an unfavourable steric interaction with the DNA (Fig. 3b). Analysis of the ternary complex shows that this short amino-terminal segment moves into the region between the primary metal-binding site and the backbone phosphates of the nucleic acid. In this new extended conformation, the main-chain oxygen of Leu 4 is now positioned within hydrogen-bonding distance of a water molecule, which is itself a ligand to the metal ion at the primary ion-binding site (Fig. 4d). Such a water-mediated interaction is presumed to be of lower binding energy than a direct bond to the metal ion, and may contribute to facile exchange of the metal ion. In the structure of the apo-repressor, the backbone carbonyl oxygen of Leu 4 is hydrogen-bonded to the backbone amide of Thr 8, forming part of the first turn of α -helix A, whereas the carbonyl oxygen of Asp 3 forms a hydrogen bond with the side chain of Thr 7. Together, the caliper-like movement of the DNA-recognition domains and the helix-to-coil transition of the N-terminal six residues account for the activation of DtxR by transition metal ions.

DtxR belongs to the HTH motif class of prokaryotic repressors. Examples of other repressors from microorganisms that use this motif for DNA recognition include the *trp* repressor^{15–17} among others¹⁰. Of these, the *trp* repressor is closest to DtxR in that binding of the co-repressor, tryptophan, causes a conformational change that results in a caliper-like domain movement which brings these motifs closer together. The fit to the major groove is thereby optimized. However, the mechanism of activation of these two repressors is very different. The two dimers of the *trp* repressor are arranged so that the dimers are interacting with each other¹⁶. In contrast, the dimers of DtxR are arranged so that they are positioned on almost opposite sides of the nucleic acid and do not interact with each other. So far, DtxR is the only prokaryotic repressor for which two dimers that are structurally independent of each other bind to a target DNA. In the *trp* repressor, the co-repressor tryptophan stabilizes the active conformation and binds at the protein–DNA interface. In contrast, in DtxR the metal ion stabilizes the dimer, induces a conformational change (caliper-like motion), and triggers a helix-to-coil transition at the N terminus. This transition in DtxR is driven by the formation of a second-sphere ligand to the activating metal ion which relieves what would otherwise be a steric block to nucleic-acid binding. Thus, the diphtheria toxin repressor has a unique mechanism of activation. □

Methods

Crystallization. Diffraction-quality single crystals of the ternary Ni(II)-DtxR(C102D)–DNA complex were obtained by the hanging-drop vapour diffusion method at room temperature. Recombinant DtxR(C102D) was overexpressed and purified as described⁹. 1 mM NiCl₂ was added to the protein solution, which was then mixed at a molar ratio of protein:DNA of 1:1.5, with a 33-bp synthetic oligonucleotide containing the corynebacteriophage β *tox* operator sequence (5'-ATATAATTAGGATAGCTTTACCTAATTATTTTA-3' and its complement 5'-TTAAATAATTAGGTAAAGCTATCCTAATTATA-3'). This protein–DNA mixture, in 100 mM Tris, pH 7.5, was then concentrated to 15 mg ml⁻¹. A 2- μ l drop of the protein–DNA solution was combined with 1 μ l reservoir solution containing 6–12% PEG4000, 10 mM MgCl₂ and 100 mM MES buffer, pH 6.5, and equilibrated against 1 ml of reservoir solution.

Crystallographic analysis. X-ray diffraction data were measured at the Stanford Synchrotron Radiation Source using a Brandeis CCD detector and a wavelength of 1.08 Å. Data obtained from three crystals were reduced, merged and scaled (Table 1) using the programs DENZO and SCALEPACK¹⁸. The symmetry of the crystals corresponds to tetragonal space group *P*₄₁ or *P*₄₃, with unit cell dimensions *a* = *b* = 117.0 Å, *c* = 145.3 Å. Two dimeric Ni(II)-DtxR(C102D) proteins, 8 Ni(II) ions, and one duplex DNA operator molecule occupy the asymmetric unit. The space group ambiguity was later resolved using molecular replacement, and is found to be *P*₄₁. A non-crystallographic

Table 1 Crystallographic analysis

Resolution (Å)	20–3
Observed reflections ($I/\sigma(I) > 0$)	107,910
Unique reflections	34,176
Signal to noise ratio ($I/\sigma(I)$)	7.6
<i>R</i> _{merge} *	0.096
Refinement	
Resolution (Å)	8–3
Reflections ($F/\sigma(F) > 0$)	32,605
Completeness (%)	88
<i>R</i> -factor ($F/\sigma(F) > 0$)	0.24
<i>R</i> -factor (highest resolution)	0.38
<i>R</i> -free value ($F/\sigma(F) > 0$)	0.28
Model	
Non-hydrogen atoms	
Protein	3,780
Metal ions	8
Nucleic acid	1,347
R.m.s. deviation from expected geometry†	
bond lengths (Å)	0.015
bond angles (deg)	1.6
dihedral angles (deg)	23
impropers (deg)	1.2
Overall <i>B</i> -value (Å ²)	37

* $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i$, where I_i is the scaled intensity of the *i*th measurement and $\langle I \rangle$ is the mean intensity for that reflection.

† R -factor = $\sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$, for all reflections used in refinement.

‡ Engh and Huber¹⁹ parameters for expected stereochemistry were used in refinement.

§ A control set of 5% of these reflections, chosen randomly, were used to calculate an *R*-free to monitor the progress of refinement²⁰.

|| The highest resolution shell of data refers to Bragg spacing values between 3.2 and 3.0 Å.

two-fold axis was found to be normal to the *c* axis, allowing us initially to reduce the data in the pseudo-space group *P*₄₁2₁2 with one Ni(II)-DtxR(C102D) dimer and half the DNA molecule in the asymmetric unit. This pseudo-symmetry facilitated solution of the structure by molecular replacement, using the program AMORE¹⁹, as implemented in the CCP4 program package²⁰. An ideal 11-mer B-form DNA model was built using QUANTA (Molecular Simulations) and used with the atomic coordinates of residues 3 to 136 of the apo-repressor² (accession number 1DPR) to solve the structure of the ternary complex in space group *P*₄₁2₁2 using data extending from 12–4 Å resolution. The positions of the protein and DNA molecules were optimized by rigid-body refinement using XPLOR²¹, yielding an *R*-factor of 0.53. In space group *P*₄₁2₁2, the 33-bp DNA molecule is positioned on a crystallographic two-fold axis parallel to the *a* axis. Because the actual DNA sequence is not an ideal palindrome, the structure was expanded in the lower-symmetry space group *P*₄₁. The orientation of the nucleic acid was assigned using simulated annealing omit maps²².

Model building and refinement. Model rebuilding was performed with the molecular graphics program O (ref. 23). Simulated annealing omit maps²² were used for some regions of the model and for the nucleic acid to help in the interpretation of the electron density. The structure was refined with XPLOR, using non-crystallographic symmetry restraints between the two dimers in the asymmetric unit, to a crystallographic *R*-factor of 0.24 and an *R*-free of 0.28 for all data with $F_{\text{obs}}/\sigma(F_{\text{obs}}) > 0$ from 8 to 3 Å resolution. Grouped temperature factor values were refined, giving an average of 37 Å² for the complex. By homology to the DNA-free holo-repressor, a single water molecule forming a sixth ligand to the co-repressor Ni(II) ion is included with the structure. The final protein–DNA model represents about 61% of the total mass of the DNA–operator complex, because residues from 121 to the C terminus of the repressor have weak, uninterpretable electron density, which is not unusual for the structure of this repressor^{2–4}. When compared with each other, subunits within each protein dimer have an r.m.s. deviation of 0.37 Å for the C α atoms. In comparison, the two dimers in the asymmetric unit have an r.m.s. deviation of 0.17 Å between each other. Analysis of the main chain (ϕ , ψ) angles²⁴ shows reasonable conformations, as all values are in allowed regions on a Ramachandran plot, except for Val 5, for which the peptide bond Leu 4–Val 5 is involved in a water-mediated interaction to the metal ion.

Refinement of the Ni-activated holo-repressor. The data of the DNA-free Ni(II)-DtxR(C102D) holo-repressor³ were re-examined. Further refinement of this complex gave an improved *R*-factor of 0.25 and an *R*-free of 0.32 ($F_{\text{obs}}/\sigma(F_{\text{obs}}) > 2$) at 2.4 Å resolution, with good stereochemistry for the model.

The atomic coordinates of this DNA-free Ni(II)-DtxR(C102D) were used to update accession entry 1TDX of the Protein Data Bank. The protein structure did not change substantially, except at the N-terminal region, where the electron density improved. Although residues 3 to 6 had originally been interpreted in a helical form³, it is now clear that these six residues are actually in an extended conformation, such that residue 4 is now found close to the primary metal-ion-binding site. Overlap of the holo-repressor with and without nucleic acid shows a r.m.s. deviation of 0.5 Å between corresponding α-carbons.

Received 12 February; accepted 27 May 1998.

1. Tao, X., Schiering, N., Zeng, H. Y., Ringe, D. & Murphy, J. R. Iron, DtxR, and the regulation of diphtheria toxin expression. *Mol. Microbiol.* **14**, 191–197 (1994).
2. Schiering, N. *et al.* Structures of the apo- and the metal ion-activated forms of the diphtheria toxin repressor from *Corynebacterium diphtheriae*. *Proc. Natl Acad. Sci. USA* **92**, 9843–9850 (1995).
3. Ding, X., Zeng, H., Schiering, N., Ringe, D. & Murphy, J. R. Identification of the primary metal ion-activation sites of the diphtheria toxin repressor by X-ray crystallography and site-directed mutational analysis. *Nature Struct. Biol.* **3**, 382–387 (1996).
4. Qiu, X. *et al.* Three-dimensional structure of the diphtheria toxin repressor in complex with divalent cation co-repressors. *Structure* **3**, 87–100 (1995).
5. Qiu, X., Pohl, E., Holmes, R. K. & Hol, W. G. High-resolution structure of the diphtheria toxin repressor complexed with cobalt and manganese reveals an SH3-like third domain and suggests a possible role of phosphate as a co-repressor. *Biochemistry* **35**, 12292–12302 (1996).
6. Somers, W. S. & Phillips, S. E. V. Crystal structure of the *met* repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature* **359**, 387–393 (1992).
7. Tao, X. & Murphy, J. R. Binding of the metalloregulatory protein DtxR to the diphtheria toxin operator requires a divalent heavy metal ion and protects the palindromic sequence from DNase I digestion. *J. Biol. Chem.* **267**, 21761–21764 (1992).
8. Must, L. M., Twiddy, E. M., Sabol, S. Z. & Holmes, R. K. in *Bacterial Protein Toxins* Vol. 6 (eds Hacker, J., Rappuoli, R., Alouf, J., Fehrenbach, F. & Feer, J.) (Fisher Stuttgart, Germany, in the press).
9. Tao, X. & Murphy, J. R. Cysteine-102 is positioned in the metal binding activation site of the *Corynebacterium diphtheriae* regulatory element DtxR. *Proc. Natl Acad. Sci. USA* **90**, 8524–8528 (1993).
10. Kercher, M. A., Lu, P. & Lewis, M. Lac repressor-operator complex. *Curr. Opin. Struct. Biol.* **7**, 76–80 (1997).
11. Albright, R. A. & Matthews, B. W. How Cro and λ-repressor distinguish between operators: The structural basis underlying a genetic switch. *Proc. Natl Acad. Sci. USA* **95**, 3431–3436 (1998).
12. Schmitt, M. P. & Holmes, R. K. Analysis of diphtheria toxin repressor-operator interactions and characterization of a mutant repressor with decreased binding activity for DNA. *Mol. Microbiol.* **9**, 173–181 (1993).
13. Wang, Z., Schmitt, M. P. & Holmes, R. K. Characterization of mutations that inactivate the diphtheria toxin repressor gene (*dtxR*). *Infect. Immun.* **62**, 1600–1608 (1994).

14. Tao, X., Zeng, H. Y. & Murphy, J. R. Transition metal ion activation of DNA binding by the diphtheria toxin repressor requires the formation of stable homodimers. *Proc. Natl Acad. Sci. USA* **92**, 6803–6807 (1995).
15. Otwinowski, Z. *et al.* Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature* **335**, 321–329 (1988).
16. Lawson, C. L. & Carey, J. Tandem binding in crystals of a *trp* repressor/operator half-site complex. *Nature* **366**, 178–182 (1993).
17. Zhang, H. *et al.* The solution structure of the *trp* repressor-operator DNA complex. *J. Mol. Biol.* **238**, 592–614 (1994).
18. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307–326 (1997).
19. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
20. Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography Version 3.1. *Acta Crystallogr. D* **50**, 760–763 (1994).
21. Brünger, A. T., Kuriyan, J. & Karplus, M. Crystallographic R-factor refinement by molecular dynamics. *Science* **235**, 458–460 (1987).
22. Hodel, A., Kim, S.-H. & Brünger, A. T. Model bias in macromolecular crystal structures. *Acta Crystallogr. A* **48**, 851–858 (1992).
23. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
24. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
25. Engh, R. A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47**, 392–400 (1991).
26. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
27. Boyd, J., Oza, M. N. & Murphy, J. R. Molecular cloning and DNA sequence analysis of a diphtheria toxin iron-dependent regulatory element (*dtxR*) from *Corynebacterium diphtheriae*. *Proc. Natl Acad. Sci. USA* **87**, 5968–5972 (1990).
28. Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R. & Sklenar, H. Conformational and helical analysis of 30ps of molecular dynamics on the d(CGCGAATTCGCG) double helix: 'Curves', dials and windows. *J. Biomol. Struct. Dynam.* **6**, 669–699 (1989).
29. Kraulis, P. J. MOLSCRIPT: A program package to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
30. Read, R. J. Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).

Acknowledgements. This work was supported by a Public Health Service Grant from the National Institute of Allergy and Infectious Diseases and (in part) by a grant from the Lucille P. Markey Charitable Trust. We thank M. Stanton and Y. Wang for help with data collection, D. Peisach and E. Peisach for rendering images, and N. Schiering, S. Chen and G. A. Petsko for discussion.

Correspondence and requests for materials should be addressed to D.R. (e-mail: ringe@binah.cc.brandeis.edu). Coordinates have been deposited in the Brookhaven Protein Data Bank under accession numbers 2tdx and 1ddn.

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

A light at the end of the tube

This assortment of products includes laser beam attenuators, motorized steppers and translators, laser confocal microscopy systems, beam analysers, as well as infrared and diode laser bars and systems.

WA-1500/WA-1000 Wavemeter

From Burleigh

Laser characterization instruments for accurate optical wavelength measurement

Designed for measuring the absolute wavelength of virtually any laser at the required accuracy, Burleigh offers systems for CW lasers, pulsed lasers and OPOs, operating from the ultraviolet through the visible and into the infrared. Absolute wavelength is measured continuously and in real-time for reporting or automatic laser control. The WA-1500 Wavemeter is stated to be an accurate, continuous, real-time laser wavelength measurement system. The WA-1500 system has most of the features of its WA-1000 counterpart, but is more affordable. These systems employ scanning Michelson, interferometer-based Wavemeter technology to determine the absolute wavelength of a test laser by comparing its interference fringe pattern with that of a built-in HeNe laser wavelength standard. The WA-1500 achieves a stated wavelength accuracy of ± 0.1 p.p.m. (± 0.1 pm at 1,000 nm); whereas the WA-1000 provides an absolute wavelength accuracy of ± 1 p.p.m. (± 1 pm at 1,000 nm). Four versions of each model are available for ultraviolet, visible, near infrared, and infrared wavelengths. Each includes a photodetector and a beamsplitter optimized for its operational wavelength range.

Reader Enquiry No. 100

LBA-300PC

From Laser Lines

Version 2.1 of this Spiricon-produced, laser beam analysis system is now available

Now running under Windows95, 98 and NT, this latest iteration includes the addition of full GPIB remote control for use in automated applications, as well improvements to the user interface, such as mouse shortcut features and improved image handling. All of the existing functions are



Keep your distance: remote control from Spiricon.

retained, including the Ultracal automatic calibration algorithm, which is said to produce improved accuracy in laser beam diameter computations, as well as in the ISO D4 σ (second moment) method for measuring beam widths.

Reader Enquiry No. 101

Eximer laser attenuators

From AG Electro-Optics

Acton Research's beam attenuators for the 193-nm ArF and 248-nm KrF excimer lasers

Eximer attenuators are multi-layer dielectric-coated, fused-silica optics that provide variation in the amount of transmitted laser energy by controlling the angle of incidence on the attenuator. For example, the 193-nm attenuator can change the output power of the ArF laser from approximately 20 per cent transmission at 0° (normal angle of incidence) to 85 % at 45° angle of incidence. The 248-nm attenuator can vary the output power of a KrF laser from approximately ten per cent of its energy output to 90 per cent output by changing the angle of incidence of the optic from normal to 45°. Applications for these attenuators include excimer laser beam attenuation, beam splitting and beam monitoring.

Reader Enquiry No. 102

Stepper Mike

From L.O.T. Oriel

Stepper motors and translators for high-resolution, 1- μ m or better, vibration-free movement in small discrete steps

Offered in 1.27- and 2.54-cm travel lengths, these units operate in full-step mode at 2 μ m per step, or in half-step mode at 1.3 and 0.7 μ m per step. This stepper should prove useful for accurate, remote control in standard micrometer applications. For sub-micrometre translation of fibre optics, laser focusing assemblies, optical mounts or other optical components, the manufacturer offers a motorized horizontal and vertical translator. Programmable single- and dual-channel controllers with RS232 and IEEE interfaces are also available.

Reader Enquiry No. 103

Modular translation stage

From New Focus

Stages with an interlocking track design

With a twist of one screw, these stages can be easily configured for multiple applications. Modular features, a low profile, a small foot-



Stage management: versatility from New Focus.

print and compatibility with standard actuators make this stage compatible with most standard actuators. Being a linear stage, it should also prove useful in general-purpose optical experiments.

Reader Enquiry No. 104

Laser systems

Diode laser bars

From Coherent

Offering 40-W bars and 30-W fibre array packaged bars at 808 nm

These bars are based on aluminium-free, active-area material and are in stock and available in a wavelength range from 800 to 820 nm. The bars are 1-cm linear arrays mounted on conduction-cooled packages. They do not require a special operating environment and are said to be easily integrated into existing applications. Achieving this difficult wavelength, says the manufacturer, demonstrates the company's ability to apply this technology widely. The fibre-array packaged bars offer up to 30 W into an 800- μ m fibre with a numerical aperture of <0.16 .

Reader Enquiry No. 105

PL2 CO₂ lasers

From Edinburgh Instruments

The company has extended the spectral range of these lasers through the use of isotopic gas fills

These gas mixtures can be used to increase the number of wavelengths obtainable with these PL2 CO₂ lasers. The PL2 laser tube is designed to allow gas to be simply and swiftly changed, making the use of isotopic gas mixes as straightforward as standard gas mixes. Using ¹³C¹⁶O₂ is said to give output powers of >3 W on the strongest lines (typically, ¹³C¹⁶O₂ has 12 lines in any of the four bands). The tunable nature of this series should prove useful in long-term studies of atmospheric conditions; and, isotopic mixtures can be chosen for coincidences with pollutant

absorption spectra, which should prove useful in the detection and monitoring of pollution. Other uses include studies of atmospheric propagation, clear-air turbulence, high-resolution spectroscopy, optical pumping and applications that require tracking and ranging.

Reader Enquiry No. 106

56IMS 633

From Melles Griot

A self-contained, temperature-stabilized diode laser system

With a long coherence length and stable output, this system has been designed for holographic interferometry, and includes automatic current and temperature-control electronics, as well as high-quality, beam-delivery optics. It is said to deliver 18-MHz, direct-modulation capability and high-power output in a small footprint, requiring only 5-VCD to operate. A front-panel C-mount is included for beam expanders and spatial filters. On the outside, a rugged, sealed enclosure keeps the laser and beam-delivery optics free from environmental contaminants. Inside the laser, there is over-voltage, reverse-voltage, transient, ESD and EMI protection. Other wavelengths, output powers and beam sizes are available upon request.

Reader Enquiry No. 107

ML3-T infrared laser

From Millennium Lasers

An infrared laser that shines in the band range 9–11 μm

This is a modified, low-power, industrial/medical laser for use in scientific applications. The ML3-T produces a TEM00 output at wavelengths selectable 9 and 11 μm with a typical output power per line of a few hundred milliwatts in CW mode. The laser head is some 45-cm long with a 10-cm cross section. It features a rugged, stabilized cavity; sealed-tube technology; frequency stability; a compact size with low weight; and a low-voltage, safety interlocking system. This laser should prove useful in environmental and remote sensing, interferometry, infrared spectrometry and injection seeding.

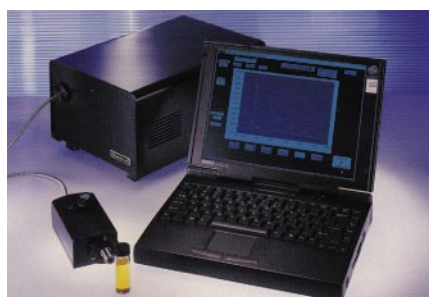
Reader Enquiry No. 108

Solution 582 and 633

From Detection Limit

Two new, low-cost Raman systems

These are dispersive CCD Raman systems: the model 633 uses an external 633-nm, 40-W HeNe laser as the excitation source, whereas the 582 system employs an internal 582-nm, 150-mW, frequency-stabilized diode laser. These fully integrated systems feature fibre-optic coupling between the laser, the remote, the miniature epi-probe



Solution 582: solves spectral analysis problems.

and the high-throughput $f/2$ imaging spectrograph. Both systems come with a Pentium-based computer controller that is loaded with mouse-driven, LabView-based DLSPEC control and data acquisition software, which includes single-button access to the embedded Grams/32 spectral analysis and presentation software.

Reader Enquiry No. 109

Confocal microscopy

MicroRadiance

From Bio-Rad

This latest instrument is said to make confocal microscopy more accessible

Featuring a miniaturized scan head, this instrument is intended to lower the price range of this technique to make it more widely accessible. MicroRadiance has a space-saving footprint, and can also be easily transferred from one microscope to another. Compatible with many inverted, standard and fixed-stage upright microscopes, the scan head needs no user realignment as all optical components are pre-set during manufacture. Motorized elements, such as apertures and filters, are all controlled from Bio-Rad's Lasersharp software. The MicroRadiance range consists of three upgradable models designed to accommodate individual budgets. The standard MR/A-1 can perform most single- and double-labelling fluorescence imaging techniques, as well as reflection and transmission imaging. The second model, MR1AG-1, has improved green/red double-labelling performance via the addition of a green He-Ne laser, which generate images with highly separated green and red channels. A further upgrade to the MR/AG-2 allows simultaneous green/red imaging through two PMT detectors.

Reader Enquiry No. 110

LSM 510

From Carl Zeiss

A compact confocal module for laser scanning microscopes with maximum spatial resolution

This fifth-generation laser scanning microscope can be combined with the high performance, fully motorized Axioplan 2 and Axiocvert microscopes to form both upright

and inverted laser microscope systems. The compact design is achieved using a high level of system integration and the shortest possible light paths to ensure high optical precision and stability with flexibility. Six detectors are integrated into an extremely small space, each has four confocal channels with individual computer-controlled, optical-spatial filter. Maximum spatial resolution is achieved through $2,048 \times 2,048$ pixels and simultaneous 12-bit AD conversion for up to 4,096 brightness levels. The LSM 510 can be integrated or retrofitted to existing motorized microscopes. It runs multi-user software through Windows NT, allowing personal application programs and routine macros to be built into the system. Functions include integration, oversampling and quasi-photon counting facilities.

Reader Enquiry No. 111

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ANNOUNCEMENT

Software reviews at www.nature.com

Turn to Nature's website where the first in a series of comparative reviews of scientific software is up and running.

The Graphical Interface

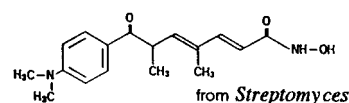
by Sharon Kardia

Nature has recruited a group of reviewers to test a variety of scientific software. The first in the series of special features has Sharon Kardia reviewing 16 graph-making packages. Each review will include a table of system functions and performance plus hyperlinks to information for contacting the software providers.

Future 'specials' will cover statistics and mathematics programs, bibliography packages and other tools of the trade.

ADVERTISEMENT

Trichostatin A



Catalog No. 204-11991

This histone deacetylase inhibitor is highly effective at low concentrations, with high specificity. For more information, please call
Wako BioProducts: (800) 992-9256; (804) 271-7677.